LEARNING SEMANTIC CORRESPONDENCES FROM NOISY DATA-TEXT PAIRS BY LOCAL-TO-GLOBAL ALIGNMENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning semantic correspondences between the structured data (e.g., slot-value pairs) and associated texts is a core problem for many downstream NLP applications, e.g., data-to-text generation. Recent neural generation methods require to use large scale training data. However, the collected data-text pairs for training are usually loosely corresponded, where texts contain additional or contradicted information compare to its paired input. In this paper, we propose a local-to-global alignment (L2GA) framework to learn semantic correspondences from loosely related data-text pairs. First, a local alignment model based on multi-instance learning is applied to build the semantic correspondences within a data-text pair. Then, a global alignment model built on top of a memory guided conditional random field (CRF) layer is designed to exploit dependencies among alignments in the entire training corpus, where the memory is used to integrate the alignment clues provided by the local alignment model. Therefore, it is capable of inducing missing alignments for text spans that are not supported by its imperfect paired input. Experiments on recent restaurant dataset show that our proposed method can improve the alignment accuracy and as a by product, our method is also applicable to induce semantically equivalent training data-text pairs for neural generation models.

1 INTRODUCTION

Learning semantic correspondences between the structured data (e.g., slot-values pairs in a meaning representation (MR)) and associated description texts is one of core problem in NLP community (Barzilay & Lapata, 2005), e.g., data-to-text generation produces texts based on the learned semantic correspondences. Recent data-to-text generation methods, especially neural-base methods which are data-hungry, adopt data-text pairs collected from web for training. Such collected corpus usually contain loosely corresponded data text pairs (Perez-Beltrachini & Gardent, 2017; Nie et al., 2019), where text spans contain information that are not supported by its imperfect structured input. Figure 1 depicts an example, where the slot-value pair `Price=Cheap` can be aligned to text span *low price range* while the text span *restaurant* doesn't supported by any slot-value pair in paired input MR. Most of previous work for learning semantic correspondences (Barzilay & Lapata, 2005; Liang et al., 2009; Kim & Mooney, 2010; Perez-Beltrachini & Lapata, 2018) focus on characterizing local interactions between every text span with a corresponded slots presented in its paired MR. Such methods cannot work directly on loosely corresponded data-text pairs, as setting is different.

In this work, we make a step towards explicit semantic correspondences (i.e., alignments) in loosely corresponded data text pairs. Compared with traditional setting, which only attempts inducing alignments for every text span with a corresponded slot presented in its paired MR. We propose a Local-to-Global Alignment (L2GA) framework, where the *local* alignment model discovers the correspondences within a single data-text pair (e.g., *low price range* is aligned with the slot `Price` in Figure 1) and a *global* alignment model exploits dependencies among alignments presented in the entire data-text pairs and therefore, is able to induce missing attributes for text spans not supported in its noisy input data (e.g., *restaurant* is aligned with the slot `EatType` in Figure 1).

Current Meaning Representation (MR)

Name	Price	Rating
Golden Palace	Cheap	low

Remaining MRs in Corpus

Name	EatType	Food
Green Man	restaurant	Italian

...

Reference

Golden Palace is a highly recommended <u>restaurant</u> specializing in <u>breakfast</u> in low price range.
Name Name   Rating   EatType   Food   Price
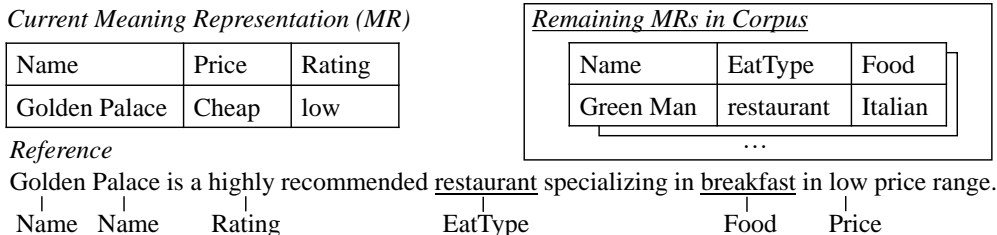
Figure 1: An example of loosely related MR-text pair. <u>underlined words</u> refer to facts that are not supported by its paired incomplete MR but has close semantics with respect to other MRs.

Specially, our proposed L2GA is composed of two parts. The *local* alignment model is a neural method optimized via a multi-instance learning paradigm (Perez-Beltrachini & Lapata, 2018) which automatically captures correspondences by maximizing the similarities between co-occurred slots and texts within a data-text pair. Our proposed *global* alignment model is a memory guided conditional random field (CRF) based sequence labeling framework. The CRF layer is able to learn dependencies among semantic labels over the entire corpus and therefore is suitable for inferring missing alignments of unsupported text spans. However, since there are no semantic labels provided for sequence labeling, we can only leverage limited supervision provided in a data-text pair. We start by generating pseudo labels using string matching heuristic between words and slots (e.g., *Golden Palace* is aligned with `Name` in Figure 1). The pseudo labels result in large portion of unmatched text spans (e.g., *low price* and *restaurant* cannot be directly matched in Figure 1), we tackle this challenge by: a) changing the calculation of prediction probability in CRF layer, where we sum probabilities over possible label sequences for unmatched text spans to allow inference on unmatched words; b) incorporating alignment results produced by the *local* alignment model as an additional memory to guide the CRF layer, therefore, the semantic correspondences captured by *local* alignment model can together work with the CRF layer to induce alignments locally and globally.

We conduct experiments of our proposed method on a recent restaurant dataset, E2E challenge benchmark (Novikova et al., 2017a), results show that our framework can improve the alignment accuracy with respect to previous methods. Moreover, our proposed method can explicitly detect unaligned errors presented in the original training corpus and provide semantically equivalent training data-text pairs for neural generation models. Experimental results also show that our proposed method can improve content consistency for neural generation models.

2 TASK FORMULATION

Here, we provide a brief description of learning alignments in loosely corresponded 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`). According to s_j, the corpus has K unique slots in total, where $K >= M$. 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. Note that for a data-text pair, not all slot-value pairs are mentioned in paired text, and not all words in text can be grounded to one of M slots in paired MR. However, some of unaligned words can be ground to one of K slots in the whole corpus. An example of alignments in a data-text pair is shown in Figure 1. The example displays differences between MR and text: (1) contradiction (`Rating:low` corresponds to the text span:*highly recommended*), (2) extra slots in text (`EatType:restaurant`). We add a special label `NULL` indicating words without any specific semantic annotation (e.g., stopwords). 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 Figure 2. It consists of two modules. The *local* model first encodes both description text X and its paired MR R using
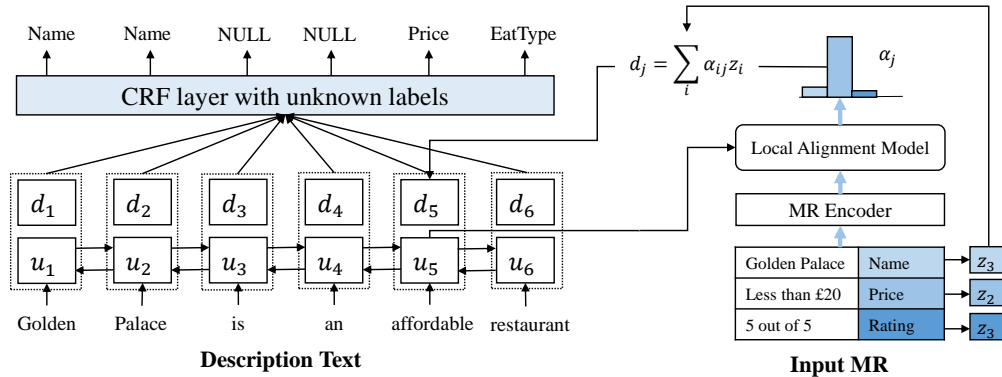
Figure 2: An illustration of Local-to-Global Alignment (L2GA). It consists of a *local* alignment model and a *global* alignment model built on a CRF layer.

contextualized encoders, then acquires semantic alignments by computing similarities in between words and slot-value pairs presented in its paired MR R. As the input MR can be incomplete, a *global* model with a specific CRF layer is proposed to exploit dependencies among alignments over the entire corpus and therefore produces possible semantic labels for text spans not supported in the paired MR. Moreover, to incorporate the alignment guidance provided by the local model, a specific memory is integrated into CRF layer to make the final alignment decision.

3.1 LOCAL ALIGNMENT MODEL

The local model tries to induce semantic labels for words in text X with respect to its paired input MR. Given a data-text pair (R, X), we can only assume that words in the description X are positively related for some slot-value pairs in R but the exact alignments are not provided. One possible way is to discover the fine-grained annotations (i.e., word alignments) from the coarse level supervisions (i.e., the similarity between a MR-text pair). Following (Perez-Beltrachini & Lapata, 2018), we formulate this task into a multi-instance learning problem (Keeler & Rumelhart, 1992). We first introduce the encoders for input MR R and description text X, then the alignment objectives to acquire the word level annotations for text X.

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

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

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

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

where E_z is a trainable slot embedding matrix.

Sentence Encoder: For description $X = (x_1, \dots, x_T)$, 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 $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_T)$.

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

Alignment Objective: Our goal is to maximize the similarity score between the MR-text pair (R, X), and we will also learn the contribution of word-level annotations for words and slot-value

pairs. Concretely, we first embed slot-value pairs in the input $R = (r_1, ..., r_M)$ into context vectors $\mathbf{c}_1, ..., \mathbf{c}_M$ using the MR encoder defined in Eq.1. Similarly, we obtain the contextual vectors $\mathbf{u}_1, ..., \mathbf{u}_T$ of description X using the sentence encoder defined in Eq.3. This similarity between MR-text pair is in turn defined on the top of the similarity scores among vector representations of slot-value pairs in R and words in description X as follows.

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

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 loss function defined in Eq.5 is to encourage related MR R and description X to achieve higher similarity than other MR $R^{'} \neq R$ and texts $X^{'} \neq X$:

$$L_{co} = \max(0, 1 - (S_{(R,X)} - S_{(R,X')})) + \max(0, 1 - (S_{(R,X)} - S_{(R',X)})) \tag{5}$$

3.2 GLOBAL ALIGNMENT MODEL

Since the data-text pairs are loosely corresponded, there exists text spans not supported by its noisy paired input. To induce semantic labels for those text spans, our proposed *global* alignment model is built on a CRF based sequence labeling framework which is capable of leveraging dependencies among alignments. Compared to conventional sequence labeling problem, our scenario differs in two aspects: i) lacking training labels for sequence labeling; ii) leveraging alignment information provided by the *local* alignment model.

To overcome the issue of lacking word-level annotations, 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 presented 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]. To allow inducing semantic labels for words with unknown types, we change the sequence paths in CRF layer. To incorporate semantic annotations learned by *local* model, particularly for text spans that are not directly recognized by string heuristics and mislabeled as an unknown entities (e.g., *affordable* in Figure 2), the alignments are treated as a soft memory to integrate into 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 W_c into the label space to compute the score P_{t,y_t} for label y_t.

$$P_t = W_c \mathbf{u}_t \tag{6}$$

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, the score of the predicted $(y_1, ..., y_T)$ as:

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

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 the only valid label sequence. However, there are entities with unknown types in our scenario (e.g., text spans *restaurant* and *affordable*

---

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

in Figure 2 are unknown entities). We instead maximize the total probability of all possible label sequences by enumerating all the possible tags for entities with unknown types. The optimization goal is defined as:

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

where Y_X refers to all possible label sequences for X, and $Y_{possible}$ contains all the possible label sequences for entities with unknown type. Note that, if there are no entities with unknown type in description text X, it is equivalent to the conventional CRF.

Integrate Local Alignment Clues: The *local* alignment model can provide alignment supervisions for words that are lexically different but semantically relevant to slot-value pairs in its paired MR. To incorporate the induced semantic labels provided by *local* alignment model, we design a specific memory into sequence labeling framework. Specially, for each word x_t in description X, we select the most probable slot s_i by computing similarity provided by *local* alignment model in Eq. 4, and compute the slot representation $\mathbf{d}_t$ as follows

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

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 Equation 6. Concretely, we modified the Eq. 6 as following:

$$P_t = W_c[\mathbf{u}_t, \mathbf{d}_t] \quad (10)$$

where $[,]$ refers to concatenation of two vectors. In this way, the alignments produced by local alignment model can act as a guidance to help inducing the 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 for path Y given the text description X. We optimize the *local* and *global* model jointly using the following training loss:

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

where L_{co} is the alignment objective of *local* alignment model defined in Eq.5 and λ is a hyper parameter and we set λ to 1 according to the validation set. 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 DATASET AND EVALUATION

Our experiments are conducted on E2E challenge (Novikova et al., 2017b) dataset, which aims at verbalizing all information from the MR. 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. Our proposed model produces alignments based on the unique slots presented in the entire dataset. The unique slots in this dataset are $\{Name, Near, EatType, Rating, Food, Price, Area, FamilyFriendly\}$.

It is difficult to evaluate the accuracy of alignment for the entire corpus, since the alignments are not provided in the original data. Due to the ambiguity of alignment boundaries (e.g., it is reasonable to tag all three words in *price is low* as `Price` or a single word *low* as `Price`), different alignment models have different alignment boundaries accordingly. Instead, alignments can be used to reproduce a refined MR by recovering slot-value pairs using the detected spans and its corresponding labels (e.g., word *price is low* and its label `Price` refers to a slot-value pair `Price:low`), more details in Appendix A.2. To make fair comparisons, we evaluate the alignments by its produced MR. The testset contains 630 unique input MRs, we randomly sample a reference for each MR, and

	Overall			Noisy data-text pairs		
	Precision	Recall	F1	Precision	Recall	F1
MIL	83.21	83.48	83.34	26.97	63.33	37.83
Distant LSTM-CRF	71.47	74.03	72.73	22.89	58.64	32.97
Modified LSTM-CRF	89.78	94.84	92.24	74.13	90.29	81.41
L2GA	**92.93**	**96.69**	**94.77**	**76.45**	**91.98**	**83.49**

Table 1: Alignment results of different methods on testset, and a subset of testset containing only loosely corresponded data-text pairs.

recruited three human annotators to label the 630 data-text pairs. The annotators were required to refine original input MRs if reference text contains contradicted or unsupported facts[2]. We calculate the precision and recall for the refined MR produced by alignment models with the annotated one.

4.2 BASELINES

We compare our proposed alignment model with the following neural baselines: i) MIL (Perez-Beltrachini & 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.

4.3 RESULTS

Table 1 presents the results of our proposed method (L2GA) with other baselines. the MIL is the *local* alignment model, which can only leverage the information within a data-text pair. Therefore it is incapable of inducing potential alignments for text spans that are not supported by its paired MR. While our proposed method L2GA can exploit dependencies among alignments globally, therefore, improves the overall alignment performance (11.43% F1 improvement with respect to MIL).

The other two methods are distant supervised sequence labeling approaches, which can be treated as simpler variations of our proposed *global* alignment model. The Distant LSTM-CRF performs worse than Modified LSTM-CRF which indicates the necessity of exploring all possible sequences in CRF layer for unknown entities. In this way, the model is able to induce a potential semantic labels for unknown type entities. Additionally, both Distant LSTM-CRF and Modified LSTM-CRF models utilize the information in its paired MR only in the creation of pseudo labels. Labels created by string matching is mislabeled as unmatched entities for text spans that are semantically equivalent but lexically different to some slot-value pairs (e.g., *afforable* is closely related to the slot-value pair `Price:Cheap` in Figure 2). While our proposed method can leverage the alignment information provided in its paired MR by the *local* alignment model simultaneously and therefore achieves substantial improvements.

As our proposed method is target on learning the alignments in loosely related data-text pairs, we pick data-text pairs in testset where the human annotated MR contains additional or contradicted slot-value pairs compared to the original MR, and we report the performance of each method on the *Noisy* data-text pairs. Results in Table 1 shows that the performance of *local* model MIL decrease dramatically, while *global* models such as Modified LSTM-CRF and L2GA are less sensitive, which proves the necessity of using a global model in learning alignments for loosely related data-text pairs. Our proposed method L2GA outperforms the Modified LSTM-CRF in both settings. The results further illustrate that both *local* and *global* models are essential for learning alignments in loosely related data-text pairs.

We report detailed alignment F1 scores of our proposed method under each slot shown in Table 2. Our proposed L2GA achieves best result in 4 out of 8 slots. The *local* model performs

---

[2]There exists a fraction 7.3% where agreements were not made, three annotators are asked to resolve the ambiguity.

	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 2: Detailed alignment results of differnt models over each slot.

	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
Local+ LSTM-CRF	87.24	**99.84**	99.51	59.04	78.51	92.47	88.69	81.10	**95.26**

Table 3: Different combinations of local models with sequence labeling framework

bad in *EatType*, which is one of the most common missing slot in the training set. The slot `familyFriendly` contains various expressions in corresponding texts, where Modified LSTM-CRF performs a lot worse than our proposed L2GA. The result indicates the necessity of integrating alignment guidance from the *local* model.

4.4 Effect of Dynamic Combination

We also investigate different ways of incorporating *local* model with the sequence labeling framework. A straight forward way is to create new pseudo labels for sequence labeling framework using the alignments produced by *local* models and train a LSTM-CRF model based on the new training labels. Table 3 gives the result. The result of the separate model performs worse than our proposed L2GA, which indicates that accurate training labels are essential to sequence labeling. L2GA dynamically integrate the results provided by the *local* model without introducing label noise for training, therefore achieves better result.

4.5 How can Alignment Help Generation?

In this section, we provide an extrinsic evaluation by testing whether alignments can help neural generation. Neural generation models trained on noisy data-text pairs suffers from *hallucination* (Reiter, 2018), where the generated texts produce contradicted or irrelevant facts with respect to its paired input. Alignments can produce a refined MR for each data-text pair, therefore, we can create a refined training corpus by applying our proposed method L2GA in training dataset. We use the

	BLEU	Err.
S2S	66.15	69.37 (374/630)
S2S+L2GA	65.21	13.33 (84/630)

Table 4: Generation results trained on different training corpus.

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 4. Vanilla S2S model trained on loosely related data-text pairs performs poorly in generation correctness. After training on the corpus refined by our proposed L2GA method, S2S model can reduce the inconsistent errors in a large margin. The results also indicates the value of studying alignments in the setting of loosely related data-text pairs, which can be of help to automatically reduce data noise in large datasets.

4.6 Qualitative Analysis

Figure 3 gives the alignment results produced by different models. We can see that *local* model MIL cannot induce the label for text spans *kid friendly* as it is contradicted with the slot-value pair `FamilyFriendly:no`. While *global* models can induce the semantic label 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 word *highly*. While L2GA

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.

can dynamically integrate the alignment results provided by the *local* model, therefore produce the semantic labels `Rating` for 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: firstly, data units are 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 & 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 & 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 & Ganti (2004) and Canisius & 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 & Knoblock, 2008; Li et al., 2009). Bellare & 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 loosely related 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. We find that our proposed method improves the alignment accuracy, and can be of help to reduce the noise in original training corpus. In the future, we will explore more challenging datasets with more complex data schema.

REFERENCES

Eugene Agichtein and Venkatesh Ganti. Mining reference tables for automatic text segmentation. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004*, pp. 20–29, 2004.

Regina Barzilay and Mirella Lapata. Modeling local coherence: An entity-based approach. In *ACL*, pp. 141–148, 2005.

Kedar Bellare and Andrew McCallum. Generalized expectation criteria for bootstrapping extractors using record-text alignment. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009, 6-7 August 2009, Singapore, A meeting of SIGDAT, a Special Interest Group of the ACL*, pp. 131–140, 2009.

Anja Belz and Eric Kow. Extracting parallel fragments from comparable corpora for data-to-text generation. In *INLG 2010 - Proceedings of the Sixth International Natural Language Generation Conference, July 7-9, 2010, Trim, Co. Meath, Ireland*, 2010.

Sander Canisius and Caroline Sporleder. Bootstrapping information extraction from field books. In *EMNLP-CoNLL 2007, Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, June 28-30, 2007, Prague, Czech Republic*, pp. 827–836, 2007.

Athanasios Giannakopoulos, Claudiu Musat, Andreea Hossmann, and Michael Baeriswyl. Unsupervised aspect term extraction with b-LSTM & CRF using automatically labelled datasets. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

Kazuma Hashimoto, Caiming Xiong, Yoshimasa Tsuruoka, and Richard Socher. A joint many-task model: Growing a neural network for multiple NLP tasks. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1923–1933, Copenhagen, Denmark, September 2017. Association for Computational Linguistics. doi: 10.18653/v1/D17-1206. URL https://www.aclweb.org/anthology/D17-1206.

Juraj Juraska, Panagiotis Karagiannis, Kevin Bowden, and Marilyn Walker. A deep ensemble model with slot alignment for sequence-to-sequence natural language generation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 152–162, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/N18-1014.

Jim Keeler and David E. Rumelhart. A self-organizing integrated segmentation and recognition neural net. In J. E. Moody, S. J. Hanson, and R. P. Lippmann (eds.), *Advances in Neural Information Processing Systems 4*, pp. 496–503. 1992.

Joohyun Kim and Raymond J. Mooney. Generative alignment and semantic parsing for learning from ambiguous supervision. In *COLING*, pp. 543–551, 2010.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williams College, Williamstown, MA, USA, June 28 - July 1, 2001*, pp. 282–289, 2001.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. 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*, pp. 260–270, 2016.

Xiao Li, Ye-Yi Wang, and Alex Acero. Extracting structured information from user queries with semi-supervised conditional random fields. In *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009, Boston, MA, USA, July 19-23, 2009*, pp. 572–579, 2009.

Percy Liang, Michael I. Jordan, and Dan Klein. Learning semantic correspondences with less supervision. In *ACL*, pp. 91–99, 2009.

Matthew Michelson and Craig A. Knoblock. Creating relational data from unstructured and ungrammatical data sources. *J. Artif. Intell. Res.*, 31:543–590, 2008.

Feng Nie, Jin-Ge Yao, Jinpeng Wang, Rong Pan, and Chin-Yew Lin. A simple recipe towards reducing hallucination in neural surface realisation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 2019. Association for Computational Linguistics.

Jekaterina Novikova, Ondřej Dušek, Amanda Cercas Curry, and Verena Rieser. Why we need new evaluation metrics for NLG. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 2241–2252, Copenhagen, Denmark, September 2017a. Association for Computational Linguistics. doi: 10.18653/v1/D17-1238. URL https://www.aclweb.org/anthology/D17-1238.

Jekaterina Novikova, Ondřej Dušek, and Verena Rieser. The E2E dataset: New challenges for end-to-end generation. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 201–206, Saarbrücken, Germany, August 2017b. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/D14-1162.

Laura Perez-Beltrachini and Claire Gardent. Analysing data-to-text generation benchmarks. In *Proceedings of the 10th International Conference on Natural Language Generation*, pp. 238–242, Santiago de Compostela, Spain, September 2017. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/W17-3537.

Laura Perez-Beltrachini and Mirella Lapata. Bootstrapping generators from noisy data. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 1516–1527, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/N18-1137.

Ehud Reiter. Hallucination in neural NLG. 2018. URL https://ehudreiter.com/2018/11/12/hallucination-in-neural-nlg/.

Jingbo Shang, Liyuan Liu, Xiaotao Gu, Xiang Ren, Teng Ren, and Jiawei Han. Learning named entity tagger using domain-specific dictionary. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

Sebastian Walter, Christina Unger, and Philipp Cimiano. A corpus-based approach for the induction of ontology lexica. In *Natural Language Processing and Information Systems - 18th International Conference on Applications of Natural Language to Information Systems, NLDB 2013, Salford, UK, June 19-21, 2013. Proceedings*, pp. 102–113, 2013.

Victor Zhong, Caiming Xiong, and Richard Socher. Global-locally self-attentive encoder for dialogue state tracking. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1458–1467, Melbourne, Australia, July 2018. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/P18-1135.

A APPENDIX

A.1 TRAINING DETAILS

For all models, we use fixed pre-trained GloVe vectors (Pennington et al., 2014) and character embeddings (Hashimoto et al., 2017). The dimensions of pre-trained word and character embeddings

are 300 and 100 respectively. The dimensions of trainable hidden units in LSTMs are all set to 400. We first pre-train our local model for 5 epochs and then train our proposed local-to-global model jointly with 10 epochs according to validation set. During training, we regularize all layers with a dropout rate of 0.1. We use stochastic gradient descent (SGD) for optimisation with learning rate 0.015. The gradient is truncated by 5.

A.2   RECOVER MR USING ALIGNMENTS

Given a MR-text pair (R, X) along with its induced alignments Y, our goal is to recover a refined MR R' by making use alignments Y. Intuitively, values for several slots belong to string values (e.g., text span *The Cricketers* with semantic label `Name`), where the value is directly recovered by the corresponding text spans (e.g., `Name:The Cricketers`). In the E2E dataset, there are two slots with a string value (i.e., `Name` and `Near`). The rest of slots use categorical values. To recover for categorical values, we apply a simple retrieval based method. Specifically, we collect the text spans with the detect labels (i.e., slots) in the training corpus with its corresponding slot-value pair presented in the MR (e.g., text span *kid friendly* with slot `FamilyFriendly`). Since the MR can be inaccurate, text spans with a specific label might have multiple referring slot-value pairs (e.g., text span *kid friendly* has two options `FamilyFriendly:yes` and `FamilyFriendly:no`). We calculate the frequency of candidate slot-value pairs, and use the most frequent one (e.g., *kid friendly* is recovered to `FamilyFriendly:yes` as it co-occurs with `FamilyFriendly:yes` a lot more than `FamilyFriendly:no`).