Mapping Client Messages to a Unified Data Model with Mixture Feature Embedding Convolutional Neural Network

Dingcheng Li*,Peini Liu†,Ming Huang‡,Yu Gu§, Yue Zhang¶,
Xiaodi Li‖, Daniel Dean§, Xiaoxi Liu†, Jingmin Xu†, Hui Lei§ and Yaoping Ruan§
* Big Data Lab, Baidu Inc, USA
† IBM Research, China
‡ Mayo Clinic, MN, USA
§ Watson Health Cloud, IBM, USA
¶ Singapore University of Technology and Design, Singapore
‖ Donghua University, China

ABSTRACT

Data mapping among different data standards in health institutes is often a necessity when data exchanges occur among different institutes. However, no matter rule-based approaches or traditional machine learning methods, none of these methods have achieved satisfactory results yet. In this work, we propose a deep learning method, mixture feature embedding convolutional neural network (MfeCNN), to convert the data mapping to a multiple classification problem. Multi-modal features were extracted from different semantic space with a medical NLP package and powerful feature embeddings were generated by MfeCNN. Classes as many as ten were classified simultaneously by a fully-connected soft-max layer based on multi-view embedding. Experimental results show that our proposed MfeCNN achieved best results than traditional state-of-the-art machine learning models and also much better results than the convolutional neural network of only using bag-of-words as inputs.

I. INTRODUCTION

Data mapping is often required when data exchange happened among different institutes using different data standards. In the health field, for historical reasons, hospitals, pharmaceutical factories, medical insurance companies, and other health-relevant industries usually have formed their own traditions in composing documents. Word-class institutes including World Health Organization, National Institute of Health and Health Level Seven International (HL7) have made great efforts in developing international standards for electronic health information that supports clinical practice and management, and delivery and evaluation of health services. Most of health institutes adopt the international standards to a certain degree. However, inconsistencies are still quite popular due to the diverse interpretations of the same standard or missing and errors made by medical staffs. Meanwhile, most of documents are unstructured or semi-structured with large volumes of free texts. Consequently, automatic mapping one data source to another becomes a hot research topic among data science community or natural language processing (NLP) community.

Data mapping and data transformation play a vital role in the information integration area. Since decades ago, as the relational database integration requirement arose, lots of research works have been focusing on the schema mapping problem[1,2]. The schema mapping problem involves automatic discovery of the mapping relationship between source and target data models. Rahm et al[3] classified these schema mapping methods as schema-only based, instance/content-based and combination approaches; schema-only based methods discover the mapping relationship by the schema meta data information[4]; instance-based mappings discover the association relationship by the real instances of data such as the word frequencies, value patterns and ranges[5]; combination approaches[6] tackle the mapping with both schema and real instances.

Recently, more researchers apply machine learning and sophisticated statistical techniques to determine instance level matching of schema elements. The approaches introduced by Doan et al[7] show that new mappings can be learned from known mappings to the target schema. Machine learning algorithms have been used to train models using known mapping and the models were applied to the new schema elements to map them to the targets[8]. The methods acquire probabilistic knowledge from examples provided by domain experts in order to train the models. The trained models in a domain can be applied to new schema mappings in the same domain[9]. Our approach is motivated by these works while we focus particularly on integrating clinical data represented by different standards, trying to associate these data with a unified data model, and in the end contributing to seamlessly exchanging the clinical data.

In this work, we propose a sophisticated machine learning model, Mixture Feature Embedding Convolutional Neural Network (MfeCNN) to tackle the task of data mapping. The innovation of our approach lies in applying deep learning method to the data mapping problem. Our data, HL7 messages from health providers, involve semi-structured data, which uses a non-XML encoding syntax based on segment format, supported by major medical information system vendors in the United States. The standards of HL7 v2 allow some custom fields and quite many fields can contain free text contents. Meanwhile, the data set we used is very unbalanced and one target category is very dominant. These characteristics of data leads to the complex mapping problem and we are trying to solve the data mapping problem by training an advanced model with many NLP features.

Basically, we convert the data mapping task to a classification task. Firstly, many relevant features were extracted with third-party tools as multi-modal inputs, including bag-of-words (BOG), part-of-speech (POS), syntax and concepts [10]. Secondly, feature embedding representations were learned with a CNN model to generate feature tensors. Thirdly, these feature tensors were fed into a multi-view based CNN model to predict the data mappings. The MfeCNN model was evaluated on data mapping from HL7 message to CommonSif and compared with baseline classification models based on Support Vector Machine (SVM) as well as other deep learning models like basic CNN. The results show that our model yields better performance than baseline models and indicate that our approach is a promising way to resolve the automatic data mapping problem and able to handle unbalance data. Although our proposal is tested with the medical field, the methodology is generic enough to handle any tasks of similar kinds.

II. Relevant work related to our model

A deep learning method MfeCNN is proposed here to handle the mapping problem of clinical data. In the input data, free texts themselves can be regarded as one type of features, namely BOG. In addition, we retrieve medical codes in the HL7 data as terms and language-relevant features including part-of-speech (POS) and syntax among those free texts. Traditional machine learning models like SVM or Conditional Random Field (CRF), just treat them as uniformed features without distinguishing their semantic categories. All those features are in fact from different semantic spaces and thus can be regarded from different modals and can be viewed differently. It is reasonable to construct a multi-modal and multi-view model for the mapping task. Before delving into our MfeCNN approach, we will give a quick survey of previous work related to multi-modal and multi-view models.

Ngiam et al [11] proposes an application of deep networks to learn features over multiple modalities. Their deep network is based on sparse restricted Boltzmann machines (RBM). Their system demonstrates the capacity of cross-modal feature learning, where better features for one modality such as video can be learned for other modality when multiple modalities (e.g., audio and video) are given at feature learning stage. This work gives us inspirations of making use of diverse modals to learn rich features from health data sets. Multiple modalities can be extracted from those data sets, including words, syntax and semantic roles of sentences and terminology codes.

Another work which brings us hunches is twin-view embedding for CNN [12]. According to the model, variable X_1 may have a twin-view embedding (tv-embedding) with regards to any X_2 if there exists a function g_1 such that $P(X_2|X_1) = g_1(f_1(X_1), X_2)$, where $(X_1, X_2) \in X_1 \times X_2$. The tv-embedding can be expressed as a function f_1. This proposal makes it possible for current data to find tv-embedding from unlabeled data and accordingly enhance the data representation. Further, the learnt tv-embedding can be integrated into supervised CNN with a compound sigmoid function as $\sigma(W\dot{r}_l(x) + V\dot{u}_l(x) + b)$. If there are multiple tv-embeddings, a summation can be added to the former equation as $\sigma(W\dot{r}_l(x) + \sum_{i=1}^{k} V^{(i)}\dot{u}_l^{(i)}(x) + b)$, which can be regarded as multi-view embedding (mv-embedding).

Our model framework share similarities with both above models. It is a multimodal model and it employs multi-view embedding for feature integration. One essential contribution comes from how we integrate features with multi-view embedding. Features in MfeCNN are from totally different semantic space or modalities including words, concepts, and syntax. Different integration functions for multi-view embedding are deployed as sigmoid functions in MfeCNN.

Usually, word embedding can be constructed with recurrent neural network (RNN) and long-short-term memory network (LSTM) to reflect the sequential relationships among words. However, those models only allow for strictly sequential information propositions. In human languages, the order and the dependencies between words and phrases are often important. A tree-LSTM [13] was proposed as a generalization of LSTMs to tree-structured network topologies to catch the dependencies between words and phrases and provide better semantic embeddings of words. In this work, tree-LSTM is employed to generate compound feature embeddings to enrich feature modalities for Mfe.

III. Neural Network Architecture

A. Mixture Feature Embedding Convolutional Neural Network

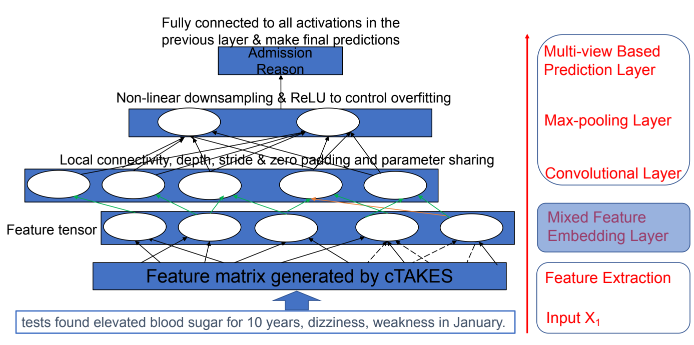

Fig. 1. Graphic View of Deep Learning Model for Data Mapping

The structure of our MfeCNN is illustrated in Figure 1. The MfeCNN model contains three main layers for feature extraction, mixture feature embedding, and deep network learning, respectively. In the first layer, a third-party tool cTAKES (clinical Text Analysis Knowledge Extraction System) [14] was

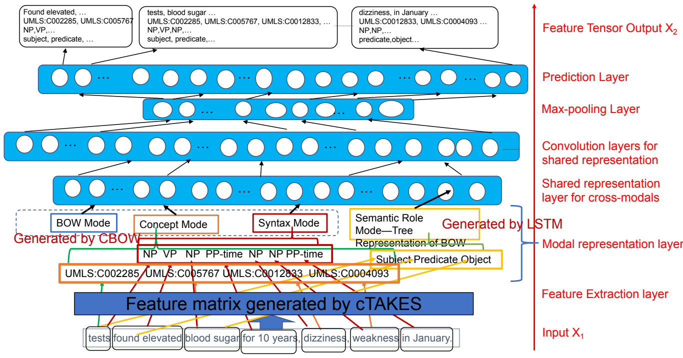

Fig. 2. Mixture Feature Embedding Layer in Details

employed to extract multi-modal features from the input sentences. The multi-modal features mainly contain Bag-of-words (BOW) modal, part-of-speech (POS) modal, concept modal and syntax modal as well as semantic role modal generated by tree-LSTM. Extracted features form a feature matrix for each sentence and are fed into Mixture Feature Embedding Layer with tensor outputs. This layer includes shared representation model for multi modalities and a convolutional neural network together with a max-pooling layer to learn feature embedding presentations as tensors. The mixture feature embedding was passed to the final deep network layer for model learning and prediction. The deep network is a convolutional neural network including a convolution lay and a max pooling layer followed by a fully-connected softmax layer based on multi-view embedding for data mapping prediction.

The main novelty in our network architecture is the inclusion of the mixture feature tensor generation layer and multi-view based prediction layer. The mixture feature embedding layer allows the network to utilize rich external resources and generate more expressive representations of features. The prediction layer with multi-view embedding can incorporate the multi-modal feature embeddings in a multi-view style and has the potential to enhance the prediction performance of data mapping. The details of the feature tensor generations and multi-view based prediction are given in next subsections.

B. Convolutional Neural Network

Convolutional neural network (CNN) here has been deployed in two different stages of our model training. Firstly, a basic CNN model was used for mixture feature embedding to generate embedding representation of feature tensor as depicted in Figure 2 and then a multi-view based CNN model is employed for the model learning of the data mapping problem as illustrated in Figure 1. Both the CNN models are a three-layer model and have a convolutional layer, a max-pooling layer and a fully-connected softmax prediction layer. The main difference between the two CNN models is the prediction layer and multi-view based CNN model has a prediction layer based on multi-view embedding.

C. Mixture Feature Embedding

Mixture feature embedding (Mfe) is the combinations of multi-modal learning and multi-view embedding as shown in Figure 2. Given the input X_1, multi-modal features can be generated by cTAKES and the extracted semantic modals mainly include POS mode, concept mode, and syntax mode. The POS mode aims at catching the property of each word in input sentences; the concept mode provides knowledge in specific domains like medicine and serve as a good discriminator of targets; the syntax mode conveys important dependency relations between words and the compositionality of a sentence usually involving phrases acts as a good indicator of the sentence nature. Due to the significance of syntax mode in human languages, further feature embedding is processed by tree-LSTM to provide rich features for semantic-role mode, which will be discussed more in the following subsection. All the generated features are firstly fed into shared representation layer for cross-modal learning and then passed to convolutional layer, max-pooling layer and prediction layer to learn feature representations as feature tensors for the MfeCNN model learning. During the model learning, Mfe, on the other hand, stretches itself to transform multi-modal learning into a multi-view embedding problem.

$$g_1(f_1(x_l), x_2) = \sigma^u(h_1 \times w^u \times r_l^u(x_1) + b^u h_1 + cx_1) \quad (1)$$

By following the idea of two-view embedding described in Equation 1 [12], Mfe attempts to find multiple two-view embeddings (multi-view embedding) for input features X_1. In addition, those multi-view embeddings are obtained from different modals or different semantic spaces and they are different from original two-view embeddings, which all are from word levels. For multi-view feature embedding from multiple semantic spaces, Equation 1 will be expanded into,

$$g_1(f_1(x_l), \mathbf{x}) = \sigma^u(\sum_{j=1}^{M} h \times w_j^u \times r_j^u(x_l) + b^u h + cx_l) \quad (2)$$

D. Enriched Features with Tree-LSTM

In the tree-LSTM, two different versions are provided, child-sum tree-LSTM and N-ary tree-LSTM and both variants have rich network topologies and can incorporate information from multiple child units. N-ary tree-LSTM takes the order and the importance of the children into consideration. Here we employ N-ary tree-LSTM to generate extra feature embeddings since the model can catch the subtle importance of each word features and this should play big roles for the training of the classification model.

In our work, constituent parser is utilized to parse a sentence into binary constituents. For example, a sentence with subject, predicate, and object may be parsed as a noun phrase (NP), the subject and a verbal phrase (VP). The VP is then parsed into a verb (V) and another NP (object). Given a constituent tree, let $C(j)$ denote the set of children of node j. The constituent tree has at most N branching factors and for each child k, a separate parameter matrix is introduced to allow N-ary tree LSTM model to learn more fine-grained conditioning on the states of a unit's children than both child-sum tree-LSTM and the flat LSTM. A sentence like above will assign the verb or VP the highest weights and the subject NP and the object NP lower weights.

E. Workable Pipeline

By combining all components discussed above, we obtained the end-to-end data mapping pipeline as illustrated in Figure 3. The pipeline is playing two stage roles, training stage and applying stage. In the training stage, the collection reader component reads the training data, including HL7 message data, target data and mapping relationship. In the applying stage, new HL7 messages go through the same pipeline and are classified by our model and each field of message contents can be mapped to the target data schema elements.

Source data in HL7 messages format will be preprocessed by an IBM parser called DFDL (Data Format Description Language) parser and are converted to a mediated format, HL7 XML. The mediated format helps us to delimitate HL7 content into segments, fields, and subfields and make it possible to analyze each part of the contents. The separated HL7 contents are analyzed by annotators of cTAKES and the CasConsumer component reads the analysis output from cTAKES. These results could be considered as NLP features of each HL7 field. These features are fed into MfeCNN and trained against the target data schema and established mappings. After the models are trained and validated, test data can go through the pipeline with similar preprocessing and feature extraction.

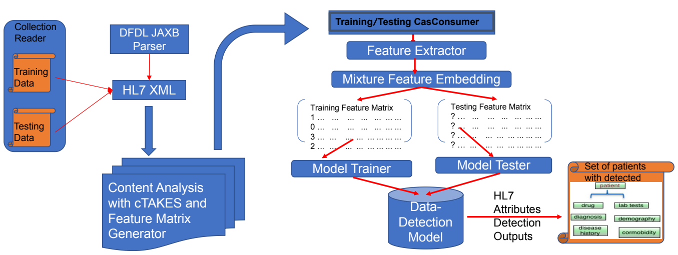

Fig. 3. End-to-End Data Mapping Pipeline of MfeCNN

IV. Experiments

In this section, we evaluate the effectiveness of our MfeCNN model on data mapping from HL7 message to CommonSif. We begin by describing the datasets used for evaluation, followed by the detailed discussion of experimental settings and final evaluation. In the evaluation, the results of MfeCNN and other models are given for comparison.

A. Datasets

Data Sets	Data Volume	Mappings
I2B2	16000	56372
Customer CDA	2400	37230
Customer HL7	1600	18604
Total	20000	112207

TABLE I
Data sets and mapping numbers

Class ID	Class Name
0	AL1.3.AllergenCode MnemonicDescription
1	DG1.4.DiagnosisDescription
2	NK1.7.ContactRole
3	NTE.3.Comment
4	OBX.5.ObservationValue
5	PID.11.PatientAddress
6	PID.5.PatientName
7	PV2.3.AdmitReason
8	Remainder
9	TXA.2.DocumentType

TABLE II
Ten target data schema elements

The input data standard for our experiments is HL7 message v2. We have totally 20000 HL7 documents with each containing only one HL7 message. We are using 3 data sets to compose these HL7 message data: I2B2 data sets[15] which contains public clinical notes for NLP research use; sample de-identified CDA data sets obtained from a hospital which records patients' medical treatment; and sample HL7 message data sets obtained from a healthcare solution provider which contains hospital charge and discharge, medical treatment and lab tests. All the data sets are converted to consistent HL7 v2 messages in advance so that we can process different format data sets with the same pipeline.

All the fields of HL7 data were manually annotated by domain experts using our target data schema CommonSIF's elements, which are considered as the ground truth. Table 1 shows our data sets in the HL7 v2 format with annotated mappings. We use about 50%, 20%, 30% of overall data sets for training, validation, and test respectively. Namely, in our experiment, we use 33662 instances to evaluate our model. In total, all instances are mapped to 10 classes (See class names listed in Tables 2) which represent 10 kinds of target data schema elements.

B. Experimental Settings

Two kinds of baselines were run to evaluate the effects of mixture feature embedding. Since our model is based on convolutional neural network, the first baseline deploys a basic CNN model with the only bag of words as inputs and the results are listed in Table III. Besides, we used SVM model to train and test data mapping as well, which enables us to make comparisons on traditional state-of-the-art machine learning model. Here both SVM and CNN baseline models use extracted features from cTAKEs as inputs and results are shown in Table IV. For the SVM model, we used libsvm[16] library to implement the classification function. Both CNN and MfeCNN were implemented with tensorflow[17] with the same network configuration as described in Section III-B. We reported standard Support, precision (Pre), recall (Rec) and F1 score for the metrics, which are defined as,

$$Support = CorrectMappingsFound \tag{3}$$

$$Pre = Support/AllMappingsFound \tag{4}$$

$$Rec = Support/AllCorrectMappings \tag{5}$$

$$F1 = 2 * Pre * Rec/(Pre + Rec) \tag{6}$$

C. Final Results

Class ID	Pre%	Rec%	F%	Support
0	46	52	48	566/1070
1	45	36	40	356/988
2	42	50	45	21/42
3	31	42	35	261/621
4	70	30	42	217/722
5	20	37	25	1316/3556
6	67	50	57	12/24
7	56	55	55	72/131
8	67	57	61	13271/23282
9	54	44	48	1420/3227
all	49.8	45.3	47.5	17512/33662

TABLE III
CNN RESULTS WITH THE ONLY BAG OF WORDS

We compared the MfeCNN results with two baseline approaches using traditional SVM model and basic CNN model with the same configurations. For all datasets we use: rectified linear units, filter windows (h) of 3, 4, 5 with dropout rate (p) of 0.5, l2 constraint (s) of 3, and mini-batch size of 50. Feature embedding dimensions vary according to the property and total vocabulary of each feature. Words themselves involve 10000 unique tokens and 200 as the feature map. Concepts involve 30000 unique identifiers and 400 as the feature map while syntax and pos have much fewer numbers (total about 100), thus embedding map only needs 20 dimensions. These values for hyperparameters were chosen via a grid search on the dev set.

Table 3 reported the baseline results of our data mapping prediction on test data conducted by the basic CNN model with the only bag of words as features. Tables 4 and 5 show the comparison of metrics of three approaches with all extracted features included. All the results are much better than that of basic CNN baseline even though the baseline is employing CNN, an advanced deep learning framework.

Although SVM, CNN, and MfeCNN all lead to very good data mappings prediction, MfeCNN achieves the best performance overall in these classes. Compared with deep learning model, SVM shows lower recalls as well as F-scores. The possible reason is that our SVM model is a linear model, which simply transforms the input to some high dimensional space to reveal the differences of classes; where deep learning model has a deep architecture with nonlinear multiple layers which combine and transform feature layers to layers, that could help to achieve better classification results. Compared with CNN, although some row like class 5 (PID.11.PatientAddress) gets lower metrics due to some default value texts, MefCNN gets much better results for most of the classes and the average F1-score is as high as 86.2%. Namely, we achieve 22% improvements than the CNN results. In addition, recently Lecun et la developed a very deep CNN model with 29 network layers to perform topic classification for ten topics by given free texts without preprocessing as inputs and the prediction accuracy achieves 73.4% [18].

These results validate that our mixture feature embedding convolutional neural network approach can indeed map the customer HL7 messages to canonical data types effectively and overcome data unbalance to some degree. Comparison with the results using traditional SVM model and CNN models shows that the combination of mixture feature embedding and convolutional neural network allows the development of sophisticated deep learning model to achieve the excellent mapping accuracy.

V. CONCLUSION AND FUTURE WORK

In this work, we implemented a novel and sophisticated deep learning framework MfeCNN for clinic data mapping. With this framework, we converted the data mapping task to a multi-label classification problem. Innovatively, we incorporate multi modalities and multi-view embedding into CNN for mixture feature tensor generation and classification prediction.

An open source tool cTAKES was utilized to perform deep language analysis for unstructured free texts so that rich linguistic features were extracted. In order to make full use of those features in the multi-modal semantic spaces, we developed a mixture feature embedding convolutional neural network to deploy those features. Mixture feature embedding realized a multi-modal and multi-view approach to digest features from different semantic spaces. This is quite different from previous approaches to do feature embedding, which usually focuses on word spaces. In contrast, we combined the feature embedding of syntax space and domain space (medical concepts) as well as word space. Our experimental results show that our approach achieves satisfactory results. In addition, the combination of mixture feature embedding and CNN plays an important role in achieving the high results.

In future work, we will focus on more features as well as improving the model as well. We may consider to use deeper network and extend MfeCNN to integrating LSTM and reinforcement learning so that more generic models can be developed for diverse data mapping tasks.

ACKNOWLEDGMENTS

This works was supported by Baidu Big Data Lab and IBM Watson Health Cloud

REFERENCES

1. J. Geller, Z. He, Y. Perl, C. P. Morrey, and J. Xu, "Rule-based support system for multiple umls semantic type assignments," *Journal of biomedical informatics*, vol. 46, no. 1, pp. 97–110, 2013.
2. H. Gu, Y. Chen, Z. He, M. Halper, and L. Chen, "Quality assurance of umls semantic type assignments using snomed ct hierarchies," *Methods of Information in Medicine*, vol. 55, no. 2, pp. 158–65, 2016.
3. E. Rahm and P. A. Bernstein, "A survey of approaches to automatic schema matching," *the VLDB Journal*, vol. 10, no. 4, pp. 334–350, 2001.
4. J. Eisenstein, B. O'Connor, N. A. Smith, and E. P. Xing, "A latent variable model for geographic lexical variation,"

Class ID	SVM Pre%	SVM Rec%	SVM F%	CNN Pre%	CNN Rec%	CNN F%	MfeCNN Pre%	MfeCNN Rec%	MfeCNN F%
0	62	88	73	65	63	63	98	68	**81**
1	75	7	14	52	46	48	72	58	**64**
2	100	25	40	59	75	66	**100**	**100**	**100**
3	48	66	55	31	82	45	**96**	**66**	**78**
4	65	17	26	97	32	48	**100**	37	**54**
5	5	30	8	26	47	33	**23**	100	37
6	**100**	**100**	**100**	97	50	65	**100**	**100**	**100**
7	40	15	21	70	23	34	**100**	**40**	**57**
8	87	89	87	96	95	95	**97**	**99**	**98**
9	96	97	96	97	97	97	**100**	**100**	**100**
Micro Ave	77.3	79.6	75.6	86.0	86.7	85.0	**89.8**	**95.9**	**89.7**
Macro Ave	67.8	52.9	59.4	69.0	59.3	63.9	**88.6**	**83.9**	**86.2**

TABLE IV
MULTI-CLASS CLASSIFICATION RESULTS FOR SVM, CNN AND MFECNN

Class ID	SVM	CNN	MfeCNN	Gold Standard
0	942	674	727	1070
1	69	454	573	988
2	11	32	42	42
3	410	509	410	621
4	122	231	267	722
5	1067	1671	3556	3556
6	24	12	24	24
7	20	30	52	131
8	20720	22117	23049	23282
9	3126	3130	3227	3227
all	26511	28860	31927	33662

TABLE V
OVERALL PERFORMANCE OF THE THREE SYSTEMS. THE VALUES ARE THE NUMBER OF TRUE POSITIVE.

in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010.

5. L. L. Yan, R. J. Miller, L. M. Haas, and R. Fagin, "Data-driven understanding and refinement of schema mappings," in *ACM SIGMOD Record*, vol. 30, no. 2. ACM, 2001, pp. 485–496.
6. C. Drumm, M. Schmitt, H.-H. Do, and E. Rahm, "Quickmig: automatic schema matching for data migration projects," in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. ACM, 2007, pp. 107–116.
7. A. Doan, P. Domingos, and A. Y. Halevy, "Reconciling schemas of disparate data sources: A machine-learning approach," in *ACM Sigmod Record*, vol. 30, no. 2. ACM, 2001, pp. 509–520.
8. B. Alexe, B. Ten Cate, P. G. Kolaitis, and W.-C. Tan, "Designing and refining schema mappings via data examples," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*. ACM, 2011, pp. 133–144.
9. J. Madhavan, P. A. Bernstein, A. Doan, and A. Halevy, "Corpus-based schema matching," in *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on*. IEEE, 2005, pp. 57–68.
10. M. Miwa and M. Bansal, "End-to-end relation extraction using lstms on sequences and tree structures," *arXiv preprint arXiv:1601.00770*, 2016.
11. J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 689–696.
12. R. Johnson and T. Zhang, "Semi-supervised convolutional neural networks for text categorization via region embedding," in *Advances in neural information processing systems*, 2015, pp. 919–927.
13. K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *arXiv preprint arXiv:1503.00075*, 2015.
14. G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, "Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications," *Journal of the American Medical Informatics Association*, vol. 17, no. 5, pp. 507–513, 2010.
15. S. N. Murphy, G. Weber, M. Mendis, V. Gainer, H. C. Chueh, S. Churchill, and I. Kohane, "Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2)," *Journal of the American Medical Informatics Association*, vol. 17, no. 2, pp. 124–130, 2010.
16. C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.
17. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.
18. A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for text classification," *arXiv preprint arXiv:1606.01781*, 2016.