

TEXTOIR: An Integrated and Visualized Platform for Text Open Intent Recognition

Hanlei Zhang^{1*}, Xiaoteng Li^{1, 2*}, Hua Xu^{1*†}, Panpan Zhang^{1, 2}, Kang Zhao^{1, 2}, Kai Gao²

¹State Key Laboratory of Intelligent Technology and Systems,

Department of Computer Science and Technology, Tsinghua University,

²School of Information Science and Engineering, Hebei University of Science and Technology

zhang-hl20@mails.tsinghua.edu.cn, xuhua@tsinghua.edu.cn

Abstract

TEXTOIR is the first integrated and visualized platform for text open intent recognition. It is composed of two main modules: open intent detection and open intent discovery. Each module integrates most of the state-of-the-art algorithms and benchmark intent datasets. It also contains an overall framework connecting the two modules in a pipeline scheme. In addition, this platform has visualized tools for data and model management, training, evaluation and analysis of the performance from different aspects. TEXTOIR provides useful toolkits and convenient visualized interfaces for each sub-module¹, and designs a framework to implement a complete process to both identify known intents and discover open intents².

1 Introduction

Analyzing user intents plays a critical role in human-machine interaction services (e.g., dialogue systems). However, many current dialogue systems are confined to recognizing user intents in closed-world scenarios, and they are limited to handle the uncertain open intents. As shown in figure 1, it is easy to identify specific purposes, such as Flight Booking and Restaurant Reservation. Nevertheless, as the user intents are varied and uncertain, pre-defined categories may be insufficient to cover all user needs. That is, there may exist some unrelated user utterances with open intents. It is valuable to distinguish these open intents from known intents, which is helpful to improve service qualities, and further discover fine-grained classes for mining potential user needs.

We divide open intent recognition (OIR) into two modules: open intent detection and open intent discovery. The first module aims to identify

User utterances	Intent Label
Book a flight from LA to Madrid.	Flight Booking
Can you get me a table at Steve's?	Restaurant reservation
Book Delta ticket Madison to Atlanta.	Flight Booking
Schedule me a table at Red Lobster.	Restaurant reservation
...	...
Can you tell me the name of this song?	Open Intent ₁
What is the calorie of this food?	Open Intent ₂

Figure 1: An example for Open Intent Recognition.

n-class known intents and detect one-class open intent (Yan et al., 2020; Lin and Xu, 2019; Shu et al., 2017). It can identify known classes but fail to discover specific open classes. The second module further groups the one-class open intent into multiple fine-grained intent-wise clusters (Vedula et al., 2020; Lin et al., 2020; Perkins and Yang, 2019). Nevertheless, the adopted clustering techniques are not able to identify known categories.

The two modules have achieved huge progress with various advanced methods on benchmark datasets. However, there still exist some issues, which bring difficulties for future research. Firstly, there are no unified and extensible interfaces to integrate various algorithms for two modules, bringing challenges for further model development. Secondly, the current methods of the two modules lack convenient visualized tools for model management, training, evaluation and result analysis. Thirdly, the two modules both have some limitations for OIR. That is, neither of them can identify known intents and discover open intents simultaneously. Therefore, OIR remains at the theoretical level, and it needs an overall framework to connect the two modules for finishing the whole process.

To address these issues, we propose TEXTOIR, the first integrated and visualized text open intent recognition platform. The platform has the follow-

* These authors contributed equally to this work.

† Hua Xu is the corresponding author.

¹Toolkit code: <https://github.com/thuiar/TEXTOIR>

²Demo code: <https://github.com/thuiar/TEXTOIR-DEMO>

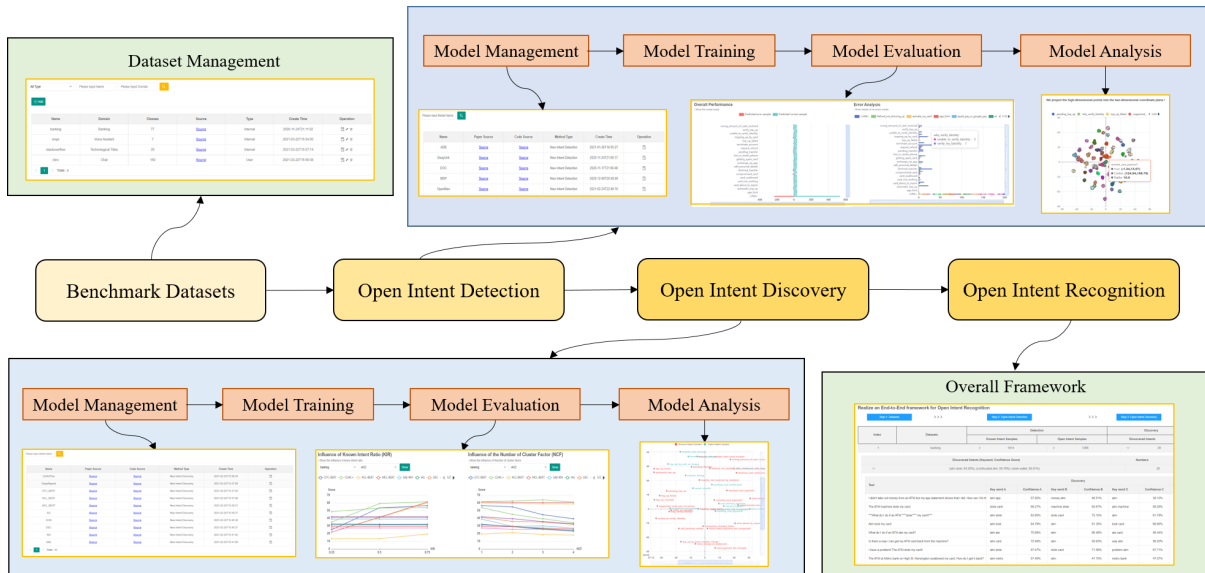


Figure 2: The architecture of the TEXTOIR platform.

ing features:

(1) It provides toolkits for open intent detection and open intent discovery, respectively. The toolkits contain flexible interfaces for data, configuration, backbone and method integration. Specifically, it integrates a series of advanced models for two modules. Each module supports a complete workflow, including data and backbone preparation with different assigned parameters, training, and evaluation. It provides standard and convenient modules to add new methods. More detailed information can be found on <https://github.com/thuiar/TEXTOIR>.

(2) It designs an overall framework combining two sub-modules naturally, achieving a complete OIR process. The overall framework integrates the advantages of two modules, which can automatically identify known intents and discover open intent clusters with recommended keywords.

(3) It provides a visualized surface for utilization. Users can leverage the provided methods or add their datasets and models for open intent recognition. We provide the front end interface for the two modules and the pipeline module. Each of the two modules supports model training, evaluation and detailed result analysis of different methods. The pipeline module leverages both the two modules and shows the complete text OIR results. More detailed information can be found on <https://github.com/thuiar/TEXTOIR-DEMO>.

2 Open Intent Recognition Platform

Figure 2 shows the architecture of the proposed TEXTOIR platform, which contains four main modules. The first module integrates a series of standard benchmark datasets. The second and third modules have toolkits for both open intent detection and open intent discovery. Besides, it visualizes the whole process (including model management, training, evaluation and result analysis) of two modules. The last module leverages the two modules in a pipeline framework to finish open intent recognition.

2.1 Data Management

Our platform supports standard benchmark datasets for intent recognition, including CLINC (Larson et al., 2019), BANKING (Casanueva et al., 2020), SNIPS (Coucke et al., 2018), and StackOverflow (Xu et al., 2015). They are all split into training, evaluation and test sets.

As shown in Figure 3, we provide unified data-processing interfaces. It supports preparing data in the format of two modules. For example, it samples known intents and labeled data with the assigned parameters for training and evaluation. Besides these labeled data, the remaining unlabeled data are also leveraged for open intent discovery. Users can see detailed statistics information from the front-end webpage and manage their datasets.

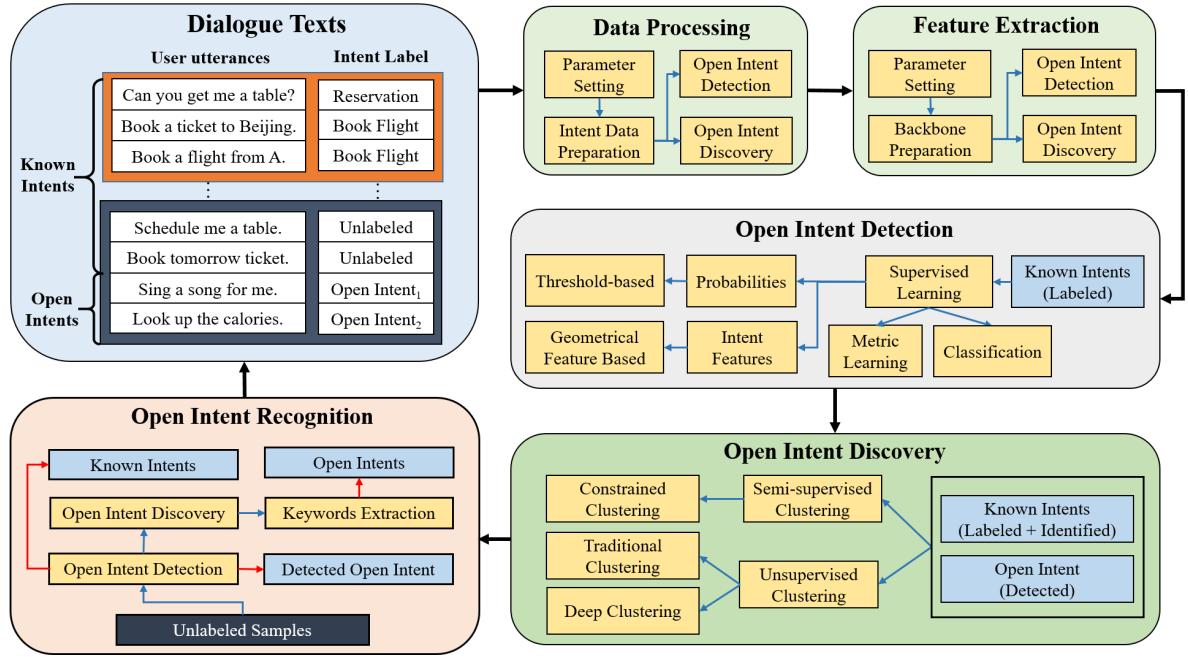


Figure 3: The architecture of Open Intent Recognition.

2.2 Models

Our platform integrates a series of advanced and competitive models for two modules, and provides toolkits with standard and flexible interfaces.

2.2.1 Open Intent Detection

This module leverages partial labeled known intent data for training. It aims to identify known intents and detect samples that do not belong to known intents. These detected samples are grouped into a single open intent class during testing. We divide the integrated methods into two categories: threshold-based and geometrical feature-based methods.

The threshold-based methods consist of MSP (Hendrycks and Gimpel, 2017), DOC (Shu et al., 2017), and OpenMax (Bendale and Boulton, 2016). These methods are first pre-trained under the supervision of the known intent classification task. Then, they leverage the probability threshold for detecting the low-confidence open intent samples. The geometrical feature-based methods include DeepUnk (Lin and Xu, 2019) and ADB (Zhang et al., 2021a). DeepUnk adopts the metric-learning method to learn discriminative intent features, and the density-based methods to detect the open intent samples as anomalies. ADB further uses the boundary loss to learn adaptive decision boundaries.

2.2.2 Open Intent Discovery

This module uses both known and open intent samples as inputs, and aims to obtain intent-wise clusters by learning from similarity properties with clustering technologies. As suggested in (Zhang et al., 2021b; Lin et al., 2020), the integrated methods are divided into two parts, including unsupervised and semi-supervised methods.

The unsupervised methods include K-Means (KM) (MacQueen et al., 1967), agglomerative clustering (AG) (Gowda and Krishna, 1978), SAE-KM, DEC (Xie et al., 2016), and DCN (Yang et al., 2017). The first two methods adopt the Glove (Pennington et al., 2014) embedding, and the last three methods leverage stacked auto-encoder to extract representations. These methods do not need any labeled data as prior knowledge and learn structured semantic-similar knowledge from unlabeled data.

The semi-supervised methods include KCL (Hsu et al., 2018), MCL (Hsu et al., 2019), DTC (Han et al., 2019), CDAC+ (Lin et al., 2020) and DeepAligned (Zhang et al., 2021b). These methods can further leverage labeled known intent data for discovering fine-grained open intents.

2.2.3 Interfaces

We provide a series of interfaces for the two modules. Firstly, the backbones are flexible and unified. For example, the primary backbone is the pre-trained BERT (Devlin et al., 2019) model, and

Realize An Overall Framework for Text Open Intent Recognition

Step 1: Datasets							Step 2: Open Intent Detection							Step 3: Open Intent Discovery						
Index	Datasets		Detection				Discovery													
			Known Intent Samples		Open Intent Samples		Discovered Intents													
1	banking		>	2093	>	783	>	19												
2	clinc		>	1275	>	1885	>	75												
Discovered Intents (Keyword, Confidence)							Numbers													
> (wrong inaccurate, 66.12%), (false true, 66.07%), (actually false, 63.36%), (right false, 62.23%), (invalid true, 62.16%)							15													
v (macaroni cheese, 62.45%), (cook apple, 51.63%), (info macaroni, 50.58%), (macaroni, 49.50%), (apple pie, 47.86%)							13													
Text				Discovery																
				Key word A	Conf A	Key word B	Conf B	Key word C	Conf C											
i want to know how nutritious an avocado typically is				nutritious avocad	88.26%	avocado typically	84.58%	avocado	78.09%											
can i substitute cream for milk				cream milk	82.39%	substitute cream	75.59%	milk	68.36%											
do you know the nutritional info for macaroni and cheese				macaroni cheese	87.09%	info macaroni	77.39%	macaroni	76.63%											
tell me the calorie content for an apple				apple	73.15%	content apple	72.22%	calorie content	53.58%											
can you tell me how i would say 'more bread please' in french				bread french	70.55%	french	53.48%	say bread	52.57%											
< 1 2 3 > Totals : 13																				

Figure 4: The pipeline framework of open intent recognition.

it supports adding new bert-based models with different downstream tasks. The open intent discovery module also supports other backbones for unsupervised clustering. Secondly, each module has the common data-loaders following the needed formats of the adopted backbones. They encode unified data vectors from the prepared data as mentioned in section 2.1. Thirdly, the parameter configurations are convenient. We extract common parameters (e.g., known intent ratio, dataset, etc.) for each module and support adding different sets of hyper-parameters for tuning each method. Finally, each approach integrates standard components of training, evaluation, and other specific functions.

3 Pipeline Framework

The two modules of open intent detection and discovery are closely related. However, there lacks an overall framework to successively invoke the two modules for both identifying known intents and discovering open intents. TEXTOIR addresses this issue with a proposed pipeline framework, as shown in Figure 3 and Figure 4.

The pipeline framework first processes the original data for two modules. Then, it feeds the labeled known intent data to the open intent detection module and trains the selected model by the users. As there is still a mass of unlabeled data containing both known and open intents, it leverages the well-trained open intent detection model to predict the unlabeled training data. The evaluated results on

training data contain identified known intents and the detected open intent. We use the predicted known intent data, detected open intent and original labeled data as the inputs of the open intent discovery module. In this case, the discovery module benefits from the detection module to obtain the augmented inputs for training. Next, the preferred clustering method selected by the users is trained to obtain the open intent clusters.

After training the two modules, they are used to perform open intent recognition on unlabeled data. Specifically, the well-trained open intent detection method is first used to predict the identified known intents and detected open intent. Then, the open intent discovery method is utilized to predict the detected open intent data to obtain the fine-grained open intent clusters. Finally, the KeyBERT toolkit (mentioned in section 4.2.2) is leveraged to extract keywords for each open intent cluster with similar-intent sentences. Therefore, our framework identifies known intents and discovers open intent samples in group with keywords as recommended labels.

4 Visualization

4.1 Training and Evaluation

Our platform provides visualized surfaces for model training and evaluation. For each method, users can change the main hyper-parameters to tune the model. When training starts, it automatically creates a record for the training process, which state

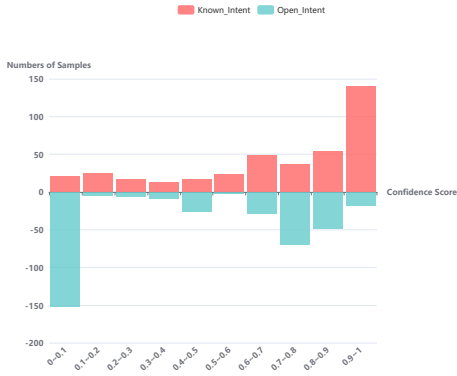


Figure 5: Known and Open intent distributions with different confidence scores.

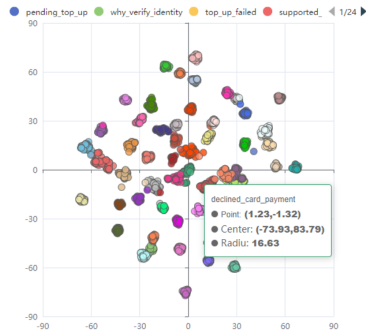


Figure 6: Visualization of the intent representations.

can be monitored by the users. When the training process finishes successfully, the trained model and related parameters are saved for further utilization.

For model evaluation, the predicted results are observed from different views. Firstly, the overall performance is shown with the number of correct and false samples for each intent class. On this basis, the number of fine-grained false-predicted classes is further shown to analyze the easily-confused intents regarding the ground truth. Secondly, the influences of the known intent ratio and labeled ratio are correspondingly shown with line charts. Users can observe the results on different selected datasets and evaluation metrics.

4.2 Result Analysis

4.2.1 Open Intent Detection

This module shows the results of identified known intent samples and detected open intent samples. For threshold-based methods, it visualizes the distribution of known and open intents with different confidence scores, which may be helpful for selecting suitable probability threshold, as shown in Figure 5.

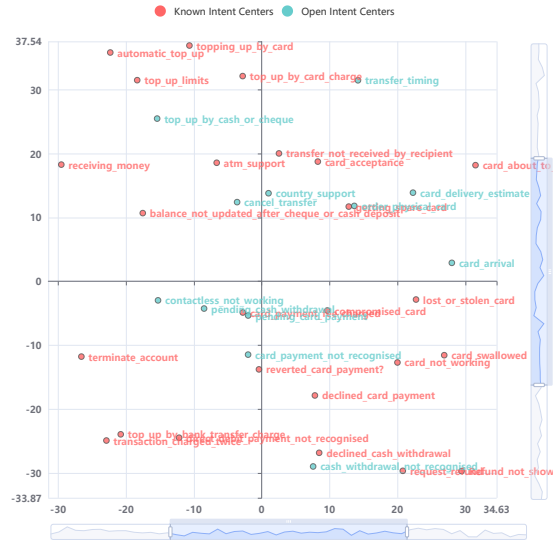


Figure 7: Intent center distribution.

For geometrical-based methods, it visualizes the intent representations on the two-dimension plane. Specifically, t-SNE (Maaten and Hinton, 2008) is applied to the high-dimension features to achieve the dimensionality reduction. Moreover, we show some auxiliary information of each point (e.g., the centre and radius of ADB), as shown in Figure 6.

4.2.2 Open Intent Discovery

For unsupervised and semi-supervised clustering methods, it shows the geometric positions of each produced cluster center with corresponding labels. These centers are categorized into the known and open classes, as shown in Figure 7. Users can mine the similarity relations of both known and open intents from observation of center distribution.

As the labels of clusters are not applicable in real scenarios, we adopt the KeyBERT³ toolkit to extract keywords for open intents in the sentence-level and cluster-level. Furthermore, it calculates the confidence scores of the keywords in the cosine similarity space. The top-3 keywords are recommended for each discovered open intent with respective confidence scores, as shown in Figure 4.

5 Experiments

We use four intent benchmark datasets mentioned in section 2.1 to verify the performance of our TEXTOIR platform. The known intent ratios are varied between 25%, 50% and 75%. The labeled proportions are varied between 50% and 100%. To evalu-

³<https://github.com/MaartenGr/keyBERT/>

ADB + DeepAligned		CLINC		BANKING		SNIPS		StackOverflow	
KIR	LR	Known	Open	Known	Open	Known	Open	Known	Open
25%	50%	89.65	86.53	84.61	63.50	87.68	32.05	82.60	45.48
25%	100%	90.88	87.71	89.08	63.67	94.79	48.89	84.13	38.87
50%	50%	91.56	87.03	84.08	69.25	94.60	61.23	80.40	55.00
50%	100%	93.42	87.80	87.50	70.61	93.83	65.84	81.73	52.37
75%	50%	91.31	86.90	83.23	68.73	95.13	63.47	79.93	48.44
75%	100%	92.80	89.21	87.89	69.83	96.10	69.11	81.24	49.78

Table 1: The open intent recognition results of ADB+DeepAligned on four datasets. "KIR" and "LR" mean the known intent ratio and labeled ratio respectively. "Known" denotes the accuracy score on known intents, and "Open" denotes the NMI score on open intents.

ate the fine-grained performance, we calculate the accuracy score (ACC) on known intents and the Normalized Mutual Information (NMI) score on open intents. We use two state-of-the-art methods of open intent detection and discovery (ADB and DeepAligned) as the components of the pipeline framework. The results are shown in Table 1.

The pipeline framework successfully connects two modules, and achieves competitive and robust results in different settings. It essentially overcomes the shortcoming of two modules, and uses the first module to identify known intents, the second module to discover open intents.

6 Related Work

6.1 Open Intent Detection

Open intent detection has attracted much attention in recent years. It aims to identify known intents while detecting the open intent. The threshold-based methods use an assigned threshold to detect the open intent. For example, MSP (Hendrycks and Gimpel, 2017) computes the softmax confidence score of each known class and regards the low-confidence samples as open. OpenMax (Bendale and Boulton, 2016) uses the Weibull distribution to produce the open class probability. (Shu et al., 2017) replaces the softmax with the sigmoid activation function and fits Gaussian distribution to the outputs for each known class. ODIN (Liang et al., 2018) adopts temperature scaling and input preprocessing technologies to obtain further discriminative probabilities for detecting open intent. The geometrical feature-based methods use the characteristics of intent features to solve this task. For example, DeepUnk (Lin and Xu, 2019) first uses the margin loss to learn the discriminative features. Then, it adopts a density-based algorithm, LOF (Breunig et al., 2000) to discover the anomaly data as the unknown intent. ADB (Zhang et al.,

2021a) learns the adaptive decision boundary for each known class among Euclidean space. However, all these methods mentioned above fail to discover fine-grained open classes.

6.2 Open Intent Discovery

Open intent discovery leverages clustering methods to help find fine-grained clusters as open intents. Unsupervised clustering methods include traditional partition-based method K-Means (MacQueen et al., 1967), hierarchical method Agglomerative Clustering (Gowda and Krishna, 1978), and density-based method (Ester et al., 1996). There are also clustering methods based on deep neural networks, such as Deep Embedded Clustering (DEC) (Xie et al., 2016), joint unsupervised learning (JULE) (Yang et al., 2016), and Deep Clustering Network (DCN) (Yang et al., 2017).

As unsupervised methods may not work well on open settings (Lin et al., 2020), researchers try to leverage some prior knowledge to improve the performance. Some methods use pairwise constraints to guide the clustering process, such as KCL (Hsu et al., 2018), MCL (Hsu et al., 2019) and CDAC+ (Lin et al., 2020). DTC (Han et al., 2019) extends DEC with temporal and ensemble information. DeepAligned (Zhang et al., 2021b) leverages clustering information to obtain aligned targets for self-supervised feature learning. However, all these clustering methods fail to identify the specific known intent classes.

7 Conclusion

We propose the first open intent recognition platform TEXTOIR, which integrates two complete modules: open intent detection and open intent discovery. It provides toolkits for each module with common interfaces and integrates multiple advanced models and benchmark datasets for the

convenience of further research. Additionally, it realizes a pipeline framework to combine the advantages of two modules. The overall framework achieves both identifying known intents and discovering open intents. A series of visualized surfaces help users to manage, train, evaluate, and analyze the performance of different methods.

Acknowledgments

This work is funded by National Key R&D Program Projects of China (Grant No: 2018YFC1707605). This work is also supported by seed fund of Tsinghua University (Department of Computer Science and Technology)-Siemens Ltd., China Joint Research Center for Industrial Intelligence and Internet of Things. We would like to thank the help from Xin Wang and Huisheng Mao, and constructive feedback from Ting-En Lin on this work.

References

- Abhijit Bendale and Terrance E Boult. 2016. Towards open set deep networks. In *Proceedings of CVPR*, pages 1563–1572.
- Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. 2000. Lof: identifying density-based local outliers. In *ACM sigmod record*, volume 29, pages 93–104. ACM.
- Iñigo Casanueva, Tadas Temcinas, Daniela Gerz, Matthew Henderson, and Ivan Vulic. 2020. Efficient intent detection with dual sentence encoders. In *Proceedings of the 2nd Workshop on NLP for ConvAI - ACL 2020*.
- Alice Coucke, Alaa Saade, Adrien Ball, Théodore Bluche, Alexandre Caulier, David Leroy, Clément Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, Maël Primet, and Joseph Dureau. 2018. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. *arXiv preprint arXiv:1805.10190*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019*.
- Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231.
- K Chidananda Gowda and G Krishna. 1978. Agglomerative clustering using the concept of mutual nearest neighbourhood. *Pattern recognition*, 10(2):105–112.
- Kai Han, Andrea Vedaldi, and Andrew Zisserman. 2019. Learning to discover novel visual categories via deep transfer clustering. In *International Conference on Computer Vision (ICCV)*.
- Dan Hendrycks and Kevin Gimpel. 2017. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *Proceedings of ICLR*.
- Yen-Chang Hsu, Zhaoyang Lv, and Zsolt Kira. 2018. Learning to cluster in order to transfer across domains and tasks. In *International Conference on Learning Representations*.
- Yen-Chang Hsu, Zhaoyang Lv, Joel Schlosser, Phillip Odom, and Zsolt Kira. 2019. Multi-class classification without multi-class labels. In *International Conference on Learning Representations*.
- Stefan Larson, Anish Mahendran, Joseph J. Peper, Christopher Clarke, Andrew Lee, Parker Hill, Jonathan K. Kummerfeld, Kevin Leach, Michael A. Laurenzano, Lingjia Tang, and Jason Mars. 2019. [An evaluation dataset for intent classification and out-of-scope prediction](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1311–1316, Hong Kong, China. Association for Computational Linguistics.
- Shiyu Liang, Yixuan Li, and R. Srikant. 2018. Enhancing the reliability of out-of-distribution image detection in neural networks. In *ICLR*.
- Ting-En Lin and Hua Xu. 2019. [Deep unknown intent detection with margin loss](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5491–5496, Florence, Italy. Association for Computational Linguistics.
- Ting-En Lin, Hua Xu, and Hanlei Zhang. 2020. Discovering new intents via constrained deep adaptive clustering with cluster refinement. In *Thirty-Fourth AAAI Conference on Artificial Intelligence*.
- Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605.
- James MacQueen et al. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA.
- Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of EMNLP 2014*, volume 14, pages 1532–1543.
- Hugh Perkins and Yi Yang. 2019. [Dialog intent induction with deep multi-view clustering](#). In *Proceedings of the 2019 Conference on Empirical Methods*

- in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4016–4025, Hong Kong, China. Association for Computational Linguistics.
- Lei Shu, Hu Xu, and Bing Liu. 2017. Doc: Deep open classification of text documents. In *Proceedings of EMNLP*, pages 2911–2916.
- Nikhita Vedula, Nedim Lipka, Pranav Maneriker, and Srinivasan Parthasarathy. 2020. Open intent extraction from natural language interactions. In *Proceedings of The WWW*, page 2009–2020.
- Junyuan Xie, Ross Girshick, and Ali Farhadi. 2016. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pages 478–487.
- Jiaming Xu, Peng Wang, Guanhua Tian, Bo Xu, Jun Zhao, Fangyuan Wang, and Hongwei Hao. 2015. [Short text clustering via convolutional neural networks](#). In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pages 62–69, Denver, Colorado. Association for Computational Linguistics.
- Guangfeng Yan, Lu Fan, Qimai Li, Han Liu, Xiaotong Zhang, Xiao-Ming Wu, and Albert Y.S. Lam. 2020. [Unknown intent detection using Gaussian mixture model with an application to zero-shot intent classification](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1050–1060, Online. Association for Computational Linguistics.
- Bo Yang, Xiao Fu, Nicholas D Sidiropoulos, and Mingyi Hong. 2017. Towards k-means-friendly spaces: Simultaneous deep learning and clustering. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3861–3870.
- Jianwei Yang, Devi Parikh, and Dhruv Batra. 2016. Joint unsupervised learning of deep representations and image clusters. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5147–5156.
- Hanlei Zhang, Hua Xu, and Ting-En Lin. 2021a. Deep open intent classification with adaptive decision boundary. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14374–14382.
- Hanlei Zhang, Hua Xu, Ting-En Lin, and Rui Lyu. 2021b. Discovering new intents with deep aligned clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14365–14373.