通用表示：

Deep contextualized word representations

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

多任务学习：

Adversarial Multi-Criteria Learning for Chinese Word Segmentation

Soft Layer-Specific Multi-Task Summarization with Entailment and Question Generation

Task Transfer by preference-Based Cost Learning

特殊网络结构：

Pointer Networks

Incorporating Copying Mechanism in Sequence-to-Sequence Learning

entence-State LSTM for Text Representation

典型sentence matching算法：

QANET: COMBINING LOCAL CONVOLUTION WITH GLOBAL SELF-ATTENTION FOR READING COMPREHENSION

Semantic Sentence Matching with Densely-connected Recurrent and Co-attentive Information

机器翻译：

Triangular Architecture for Rare Language Translation

NER：

Learning Named Entity Tagger using Domain-specific dictionary

语义单元：

Semantic-unit-based dilated convolution for multi-label text classification

Universal language model fine-tuning for text classification

Learning word vectors with linear constrains: a matrix factorization approach

其他：

Knowledgeable Reader: Enhancing Cloze-Style Reading Comprehension with External Commonsense Knowledge.