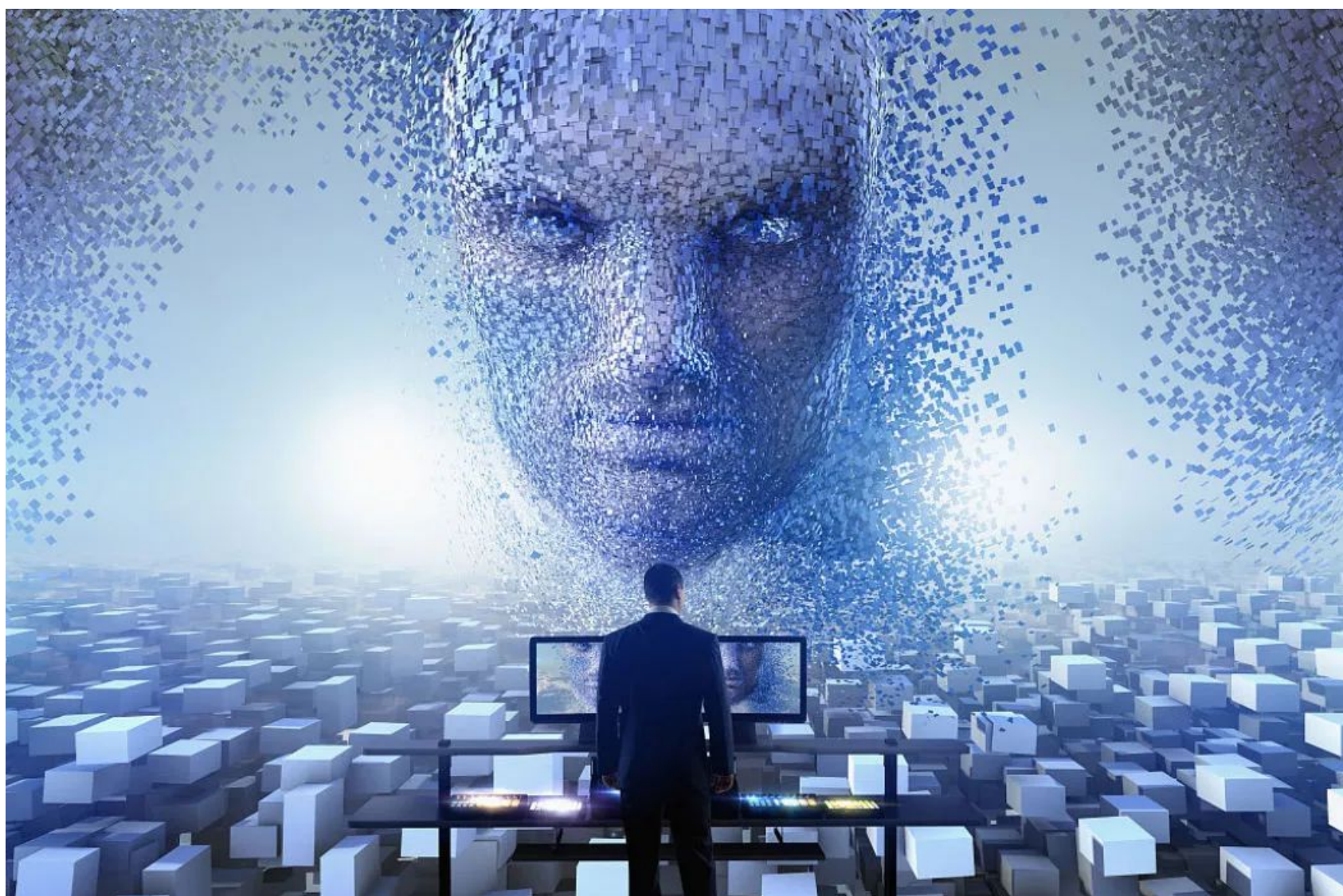


# 史上最全！阿里智能人机交互的核心技术解析

阿里妹导读：过去20多年，互联网及移动互联网将人类带到了一个全新的时代，如果用一个词来总结和概括这个时代的话，“连接”这个词再合适不过了。这个时代主要建立了四种连接：第一，人和商品的连接；第二，人和人的连接；第三，人和信息的连接；第四，人和设备的连接。



“连接”本身不是目的，它只是为“交互”建立了通道。在人机交互（Human-Computer Interaction）中，人通过输入设备给机器输入相关信号，这些信号包括语音、文本、图像、触控等中的一种模态或多种模态，机器通过输出或显示设备给人提供相关反馈信号。“连接”为“交互”双方架起了桥梁。

“交互”的演进方向是更加自然、高效、友好和智能。对人来说，采用自然

语言与机器进行智能对话交互是最自然的交互方式之一，但这条路上充满了各种挑战。如何让机器理解人类复杂的自然语言？如何对用户的提问给出精准的答案而不是一堆候选？如何更加友好地与用户闲聊而不是答非所问？如何管理复杂的多轮对话状态和对话上下文？在阿里巴巴，我们从2014年初开始对智能对话交互进行探索和实践创新，研发成果逐步大规模应用在了智能客服（针对阿里巴巴生态内部企业的阿里小蜜、针对阿里零售平台上的千万商家的店小蜜，以及针对阿里之外企业及政府的云小蜜等）和各种设备（如YunOS手机、天猫魔盒、互联网汽车等）上。

本文将对阿里巴巴在智能对话交互技术上的实践和创新进行系统的介绍。首先简要介绍智能对话交互框架和主要任务；接下来详细介绍自然语言理解、智能问答、智能聊天和对话管理等核心技术；然后介绍阿里巴巴的智能对话交互产品；最后是总结和思考。强烈建议收藏细看！

本文作者：孙健，李永彬，陈海青，邱明辉

## 1 智能对话交互框架

典型的智能对话交互框架如图1所示。其中，语音识别模块和文本转语音模块为可选模块，比如在某些场景下用户用文本输入，系统也用文本回复。自然语言理解和对话管理是其中的核心模块，广义的自然语言理解模块包括对任务类、问答类和闲聊类用户输入的理解，但在深度学习兴起后，大量端到端（End-to-End）的方法涌现出来，问答和聊天的很多模型都是端到端训练和部署的，所以本文中的自然语言理解狭义的单指任务类用户输入的语义理解。在图2所示的智能对话交互核心功能模块中，自然语言理解和对话管理之外，智能问答用来完成问答类任务，智能聊天用来完成闲聊类任务。在对外输出层，我们提供了SaaS平台、PaaS平台和BotFramework三种方式，其中Bot Framework为用户提供了定制智能助理的平台。

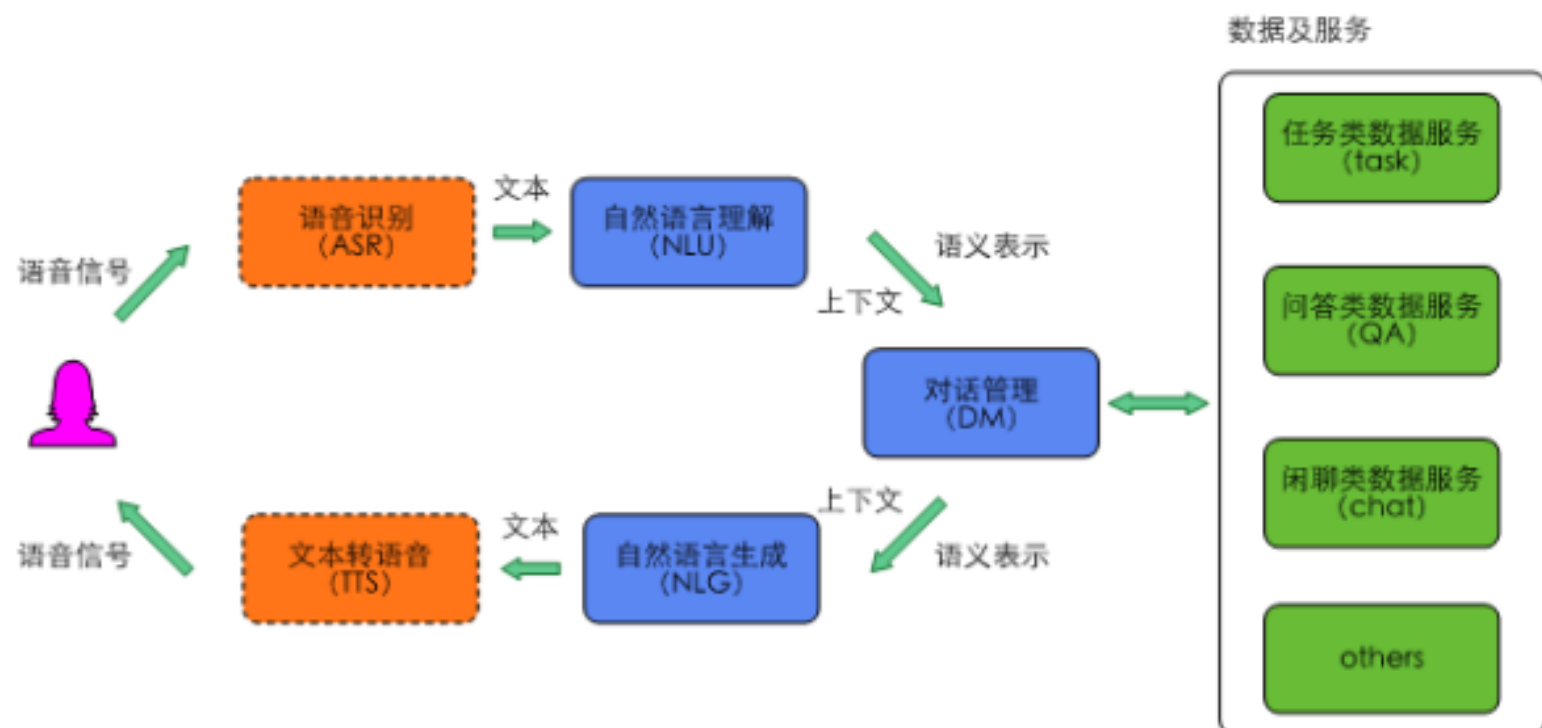


图1 智能对话交互框架

## 2 智能对话交互核心技术

智能对话交互中的核心功能模块如图2所示，本部分详细介绍智能对话交互中除输出层外的自然语言理解、智能问答、智能聊天和对话管理四个核心模块。

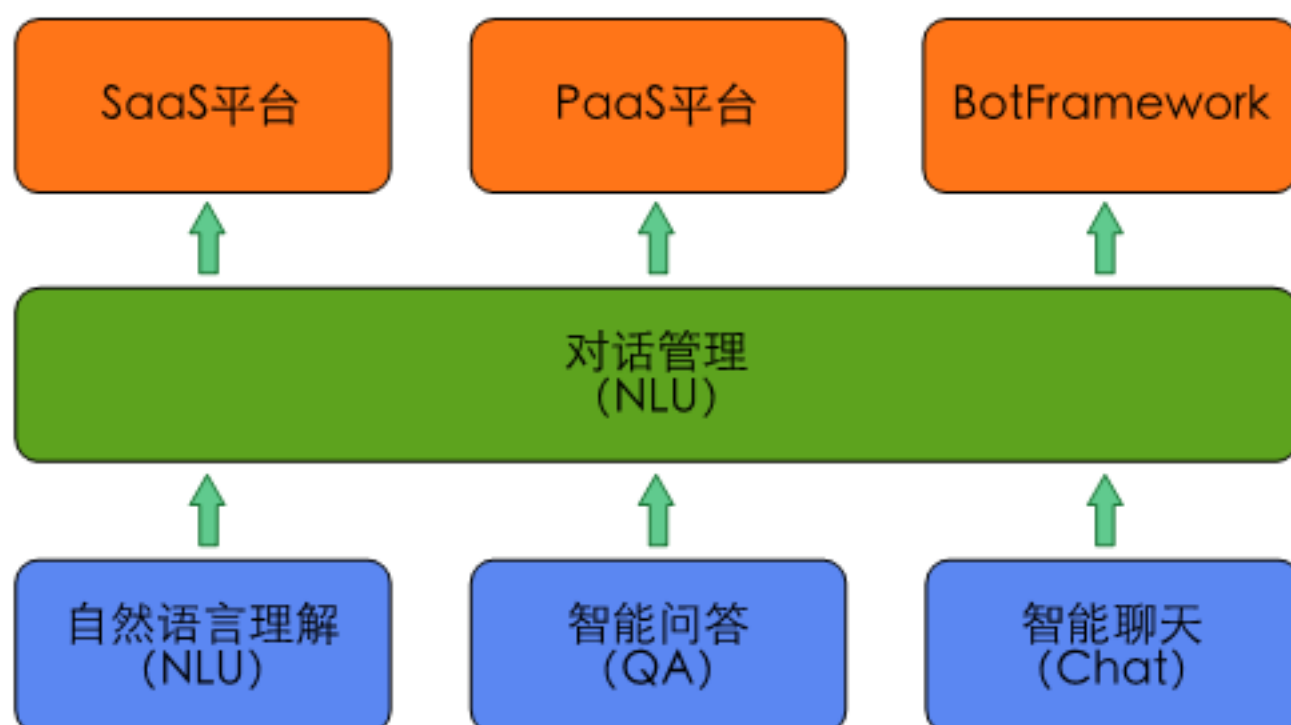


图2 智能对话交互中的核心功能模块

2.1自然语言理解

自然语言理解是人工智能的AI-Hard问题[1]，也是目前智能对话交互的核心难题。机器要理解自然语言，主要面临如下的5个挑战。

- (1) 语言的多样性
- (2) 语言的多义性
- (3) 语言的表达错误
- (4) 语言的知识依赖
- (5) 语言的上下文

表1 上下文示例

U：上海明天的天气 A：上海明天天气..... U：后天呢	U：那你嫁给我吧  A：我妈说我还小  U：我问过你妈了她说同意你嫁给我
继续延续问天气	如何正确的把闲聊接下去

注：U指用户（user），A指智能体（agent）。下同。

整个自然语言理解围绕着如何解决以上难点问题展开。

2.1.1自然语言理解语义表示



自然语言理解的语义表示主要有三种方式[2]。

- (1) 分布语义表示 (Distributional semantics)
- (2) 框架语义表示 (Frame semantics)
- (3) 模型论语义表示 (Model-theoretic semantics)

在智能对话交互中，自然语言理解一般采用framesemantics表示的一种变形，即采用领域 (domain)、意图 (intent) 和属性槽 (slots) 来表示语义结果，如图3所示。

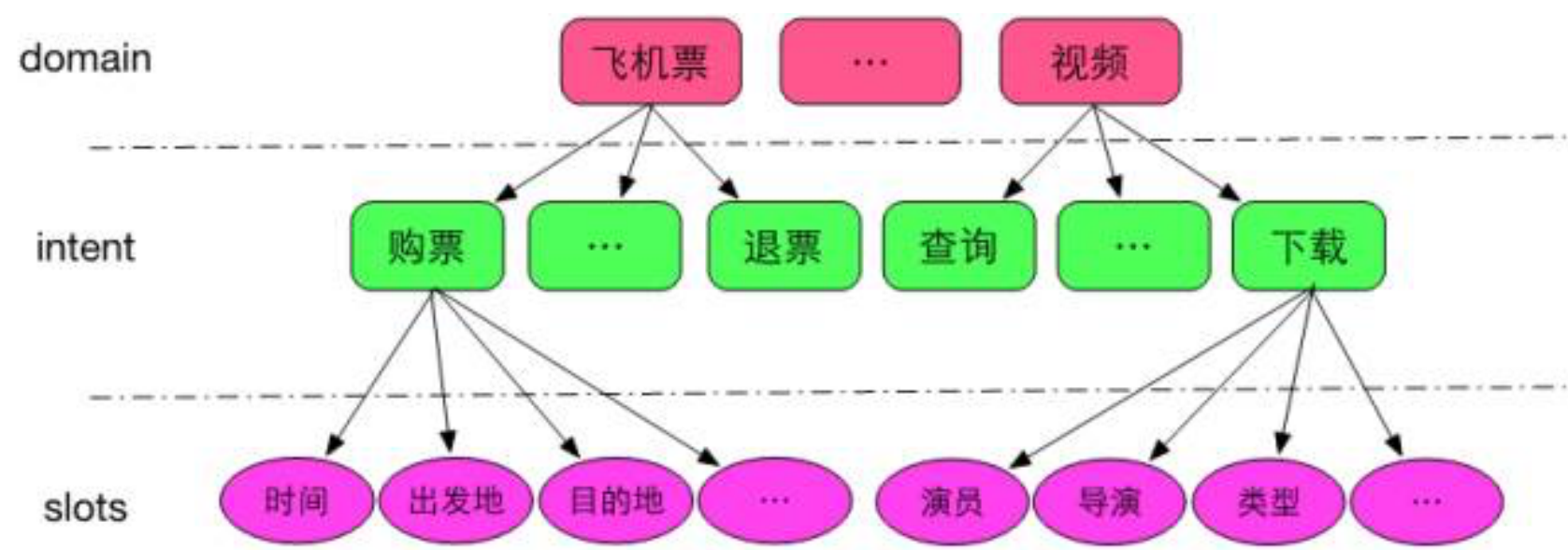


图3domain ontology示意图

在定义了上述的domain ontology结构后，整个算法流程如图4所示。

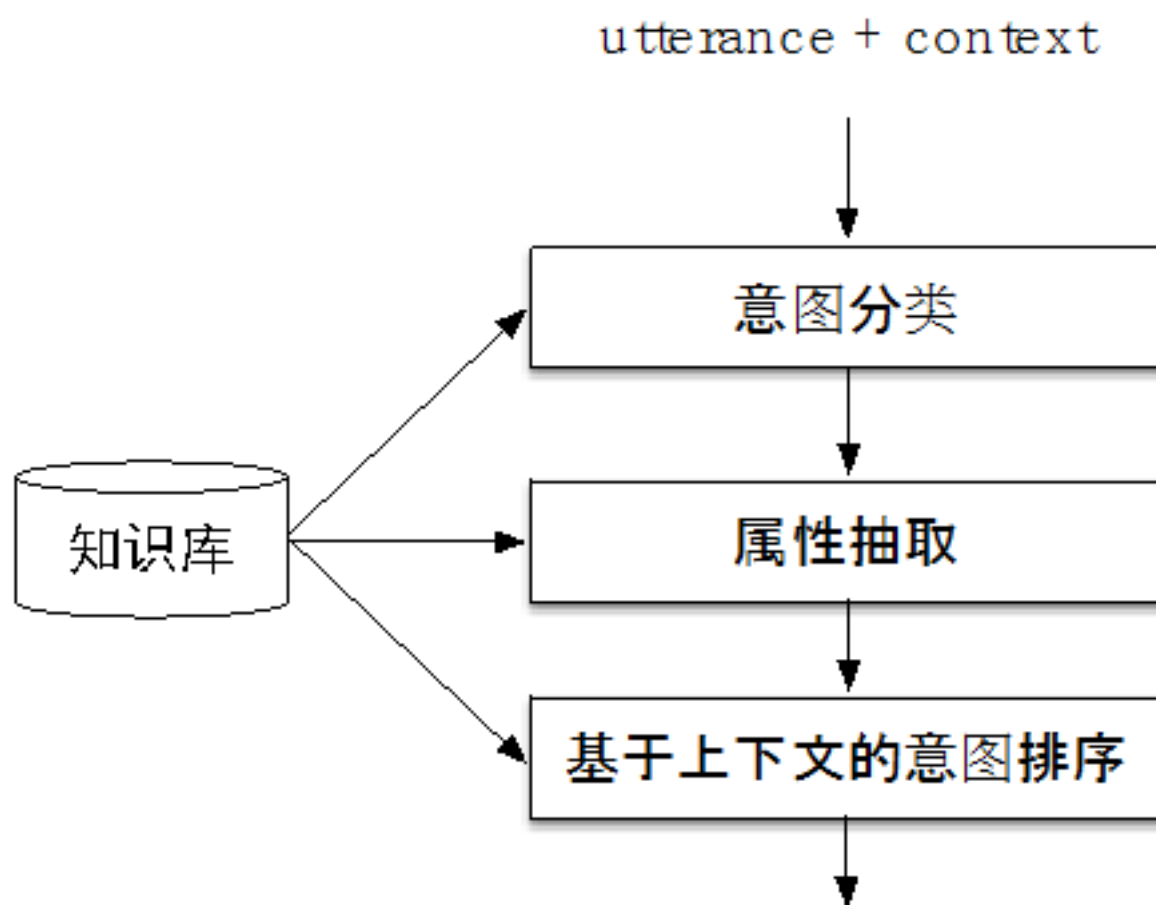


图4 自然语言理解流程简图

### 2.1.2意图分类

意图分类是一种文本分类，主要分为基于规则的方法、基于传统机器学习的方法和基于深度学习的方法，如CNN [3]、LSTM [4]、RCNN [5]、C-LSTM [6]及FastText[7]等。针对CNN、LSTM、RCNN、C-LSTM四种典型的模型框架，我们在14个领域的数据集上进行训练，在4万左右规模的测试集上进行测试，采用Micro F1作为度量指标（注：此处的训练和测试中，神经网络的输入只包含word embedding，没有融合符号表示），结果如图5所示，其中Yoon Kim在2014年提出的基于CNN[3]的分类算法效果最好。

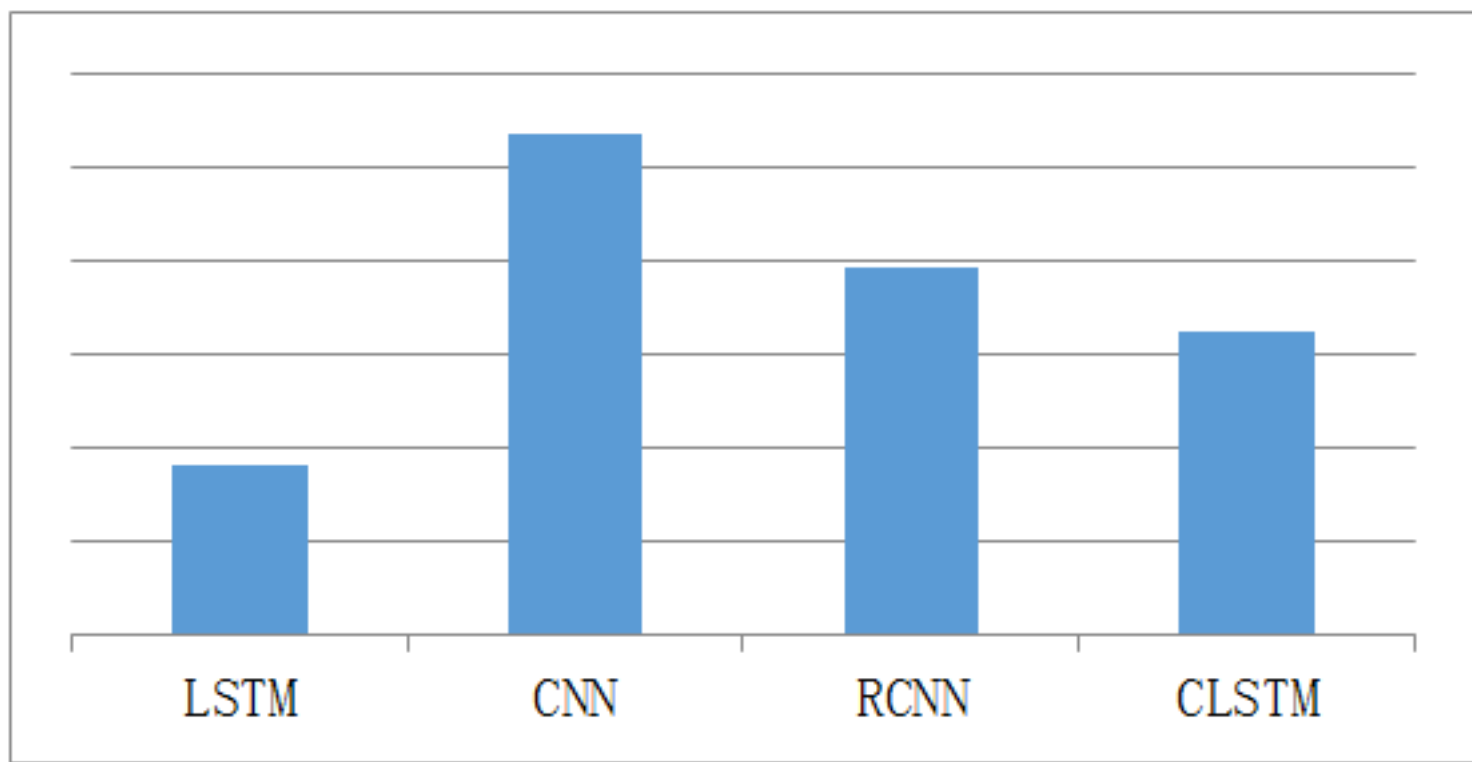


图5 四种模型的分类效果对比

单纯以word vector为输入的CNN分类效果，在某些领域上无法超越复杂特征工程的SVM分类器。如何进一步提升深度学习的效果，其中一个探索方向就是试图把分布式表示和符号表示进行融合。比如对于“刘德华的忘情水”这句话，通过知识库可以标注刘德华为singer、忘情水为song，期望能把singer和song这样的符号表示融入到网络中去。具体融合方法，既可以把符号标签进行embedding，然后把embedding后的vector拼接到wordvector后进行分类，也可以直接用multi-hot的方式拼接到word vector后面。分布式表示和符号表示融合后的CNN结构如图6所示。

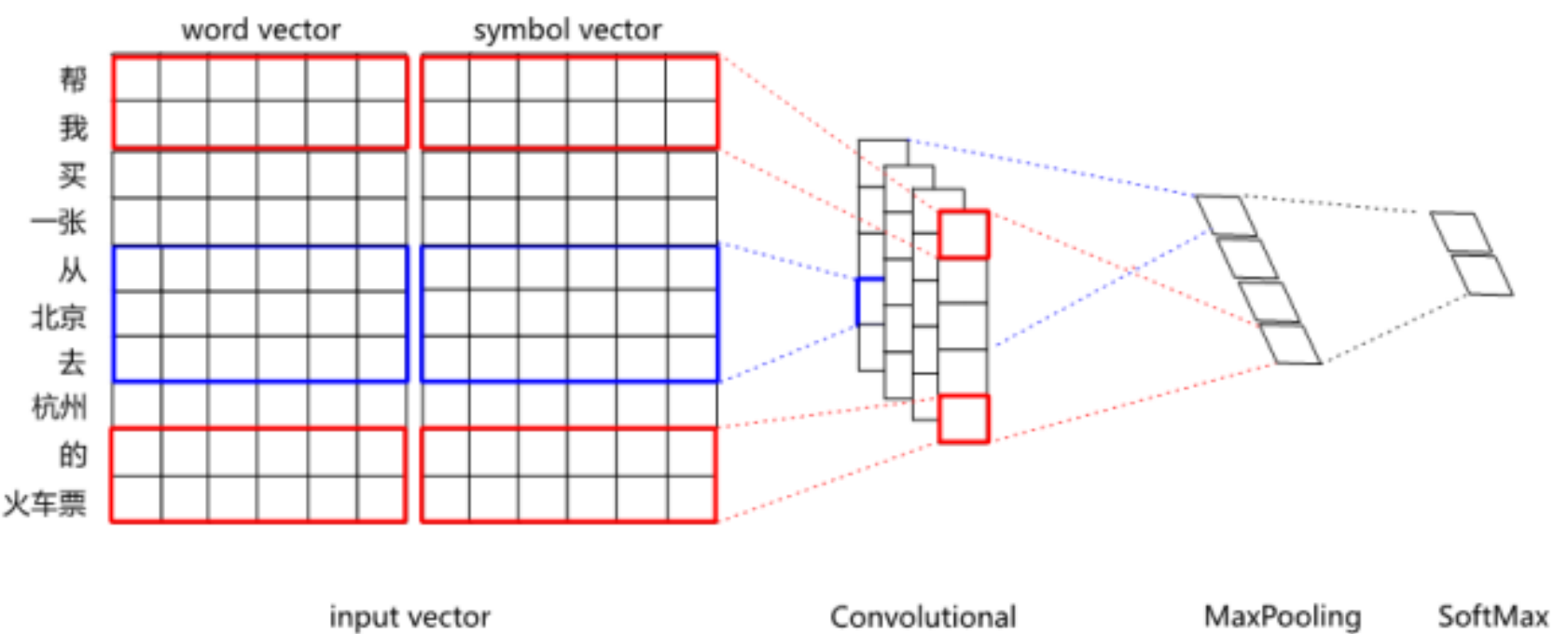


图6 分布式表示和符号表示融合后的CNN分类网络结构

经过融合后，在14个领域约4万条测试数据集上，对比融合前后的F1值（如图7所示），从中可以看出，像餐厅、酒店、音乐等命名实体多且命名形式自由的领域，效果提升非常明显。

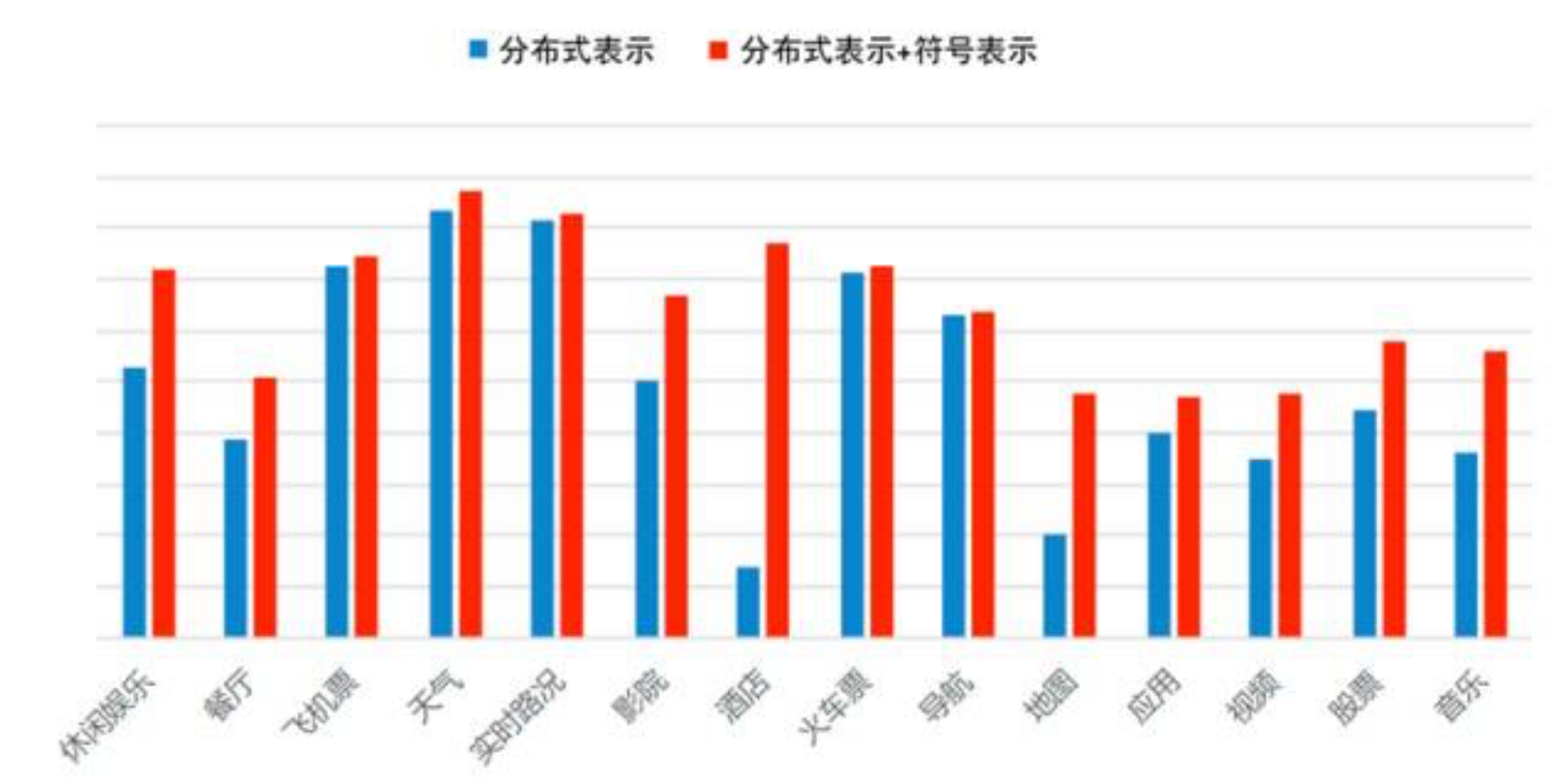


图7 在CNN中分布式表示融合符号表示前后效果对比

在以词为输入单位的CNN中，经常会遇到OOV（Out-Of-Vocabulary）问题，一般情况下会使用一个特殊向量（比如固定的随机向量或者已知词向量的平均值）来表示所有的OOV，这样做的效果肯定不够好。在我们的实现中，引入了FastText [8]来训练word vector，对于OOV，可以用其subword向量计算得到，有效地解决了OOV的问题。

在效果优化方面，除了本文中所述的word vector的动态训练和dropout之外，通过对训练数据进行数据增强（data augmentation），效果会有较大的提升。



### 2.1.3属性抽取

属性抽取问题可以抽象为一个序列标注问题，可以以字为单位进行序列标注，也可以以词为单位进行序列标注，如图8所示为以词为单位进行序列标注的示例。在这个例子中包含departure、destination和time三个待标注标签；B表示一个待标注标签的起始词；I表示一个待标注标签的非起始词，O表示非待标注标签词。

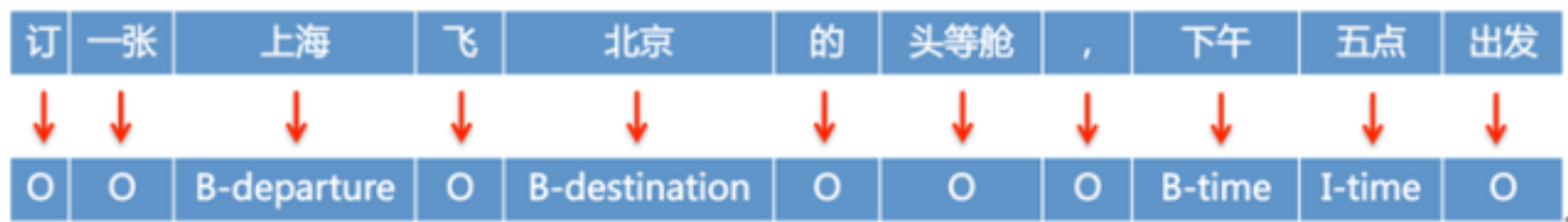


图8 序列标注示例

属性抽取的方法，包括基于规则的方法，基于传统统计模型的方法，经典的如CRF[9]，以及基于深度学习模型的方法。2014年，在ARTIS数据集上，RNN[10]模型的效果超过了CRF。此后，R-CRF [11]、LSTM[12]、Bi-RNN[13]、Bi-LSTM-CRF[14]等各种模型陆续出来。

在属性抽取这个任务中，我们采用了如图9的网络结构，该结构具有以下优点。

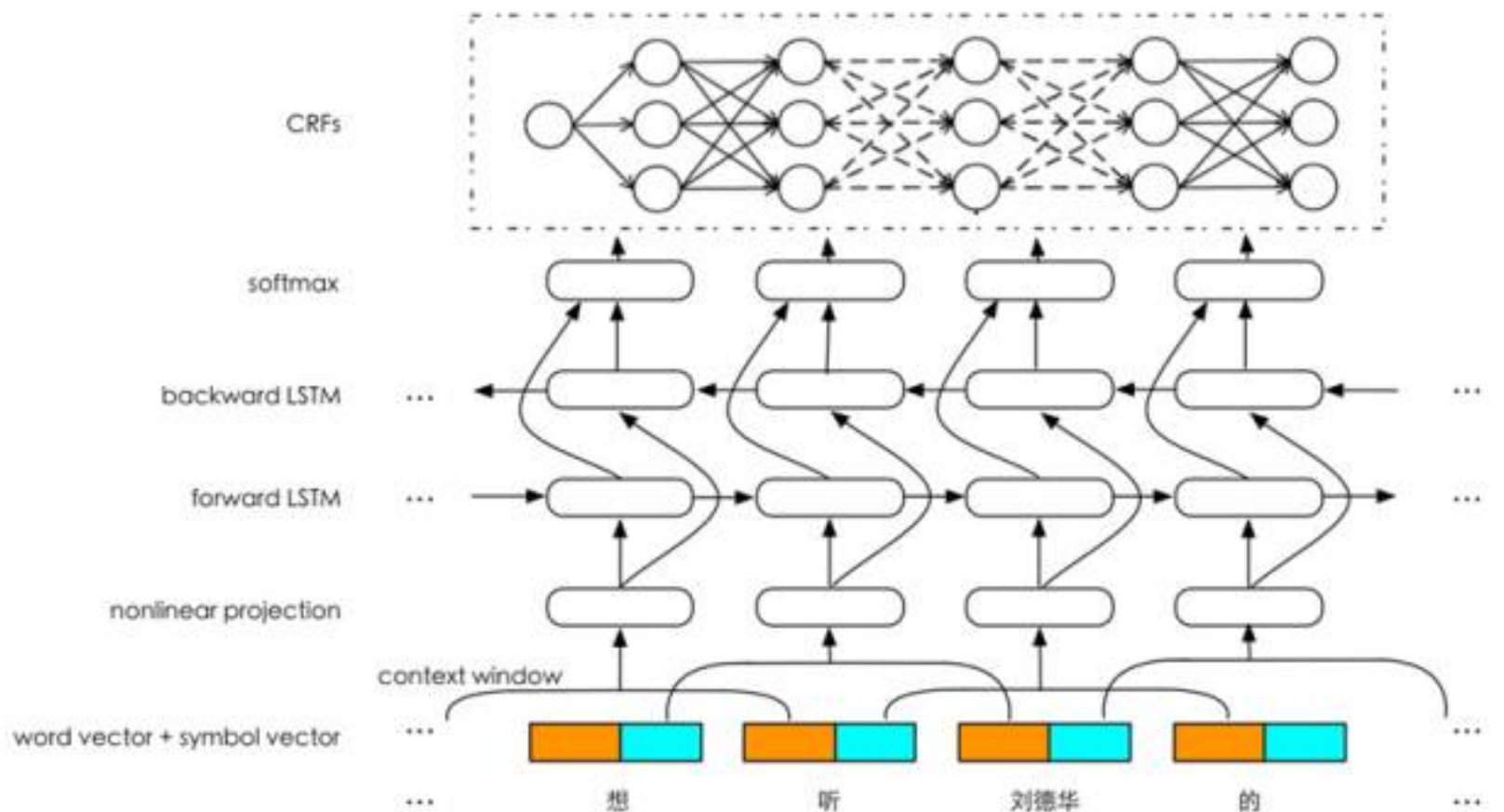


图9 属性抽取网络结构

### (1) 输入层

在输入层，我们做了三部分工作：① 采用了分布式表示（word vector）和符号表示（symbol vector）融合的方式，有效利用了分布式的上下文学习能力和符号的抽象知识表示能力；② 采用了局部上下文窗口（local context window），将窗口内的词的表示拼接在一起送入一个非线性映射层，非线性映射具有特征学习和特征降维的作用；③ 采用了FastText [8]进行word embedding的学习，可以有效解决OOV（Out-Of-Vocabulary）的问题。

### (2) Bi-LSTM层

在中间的隐藏层，采用Bi-LSTM进行特征学习，既能捕捉上文特征，也能捕捉下文特征。

### (3) 输出层

在输出层有几种典型的做法，比如Bi-LSTM+Softmax、Bi-LSTM+CRF等，Bi-LSTM+Softmax是把属性抽取在输出层当成了一个分类问题，得到

的标注结果是局部最优，Bi-LSTM+CRF在输出层会综合句子层面的信息得到全局最优结果。

## 2.1.4意图排序

在表1中，我们展示了一个例子，如果不看上下文，无法确定“后天呢”的意图。为了解决这个问题，在系统中我们设计了意图排序模块，其流程如图10所示。对于用户输入的utterance，一方面先利用分类抽取模型去判定意图并做抽取；另一方面，直接继承上文的意图，然后根据这个意图做属性抽取。这两个结果通过特征抽取后一起送入一个LR分类器，以判定当前utterance是应该继承上文的意图，还是遵循分类器分类的意图。如果是继承上文意图，那么可以把这个意图及其属性抽取结果作为最终结果输出；如果是遵循分类器分类的结果，那么可以把各个结果按照分类器分类的置信度排序输出。

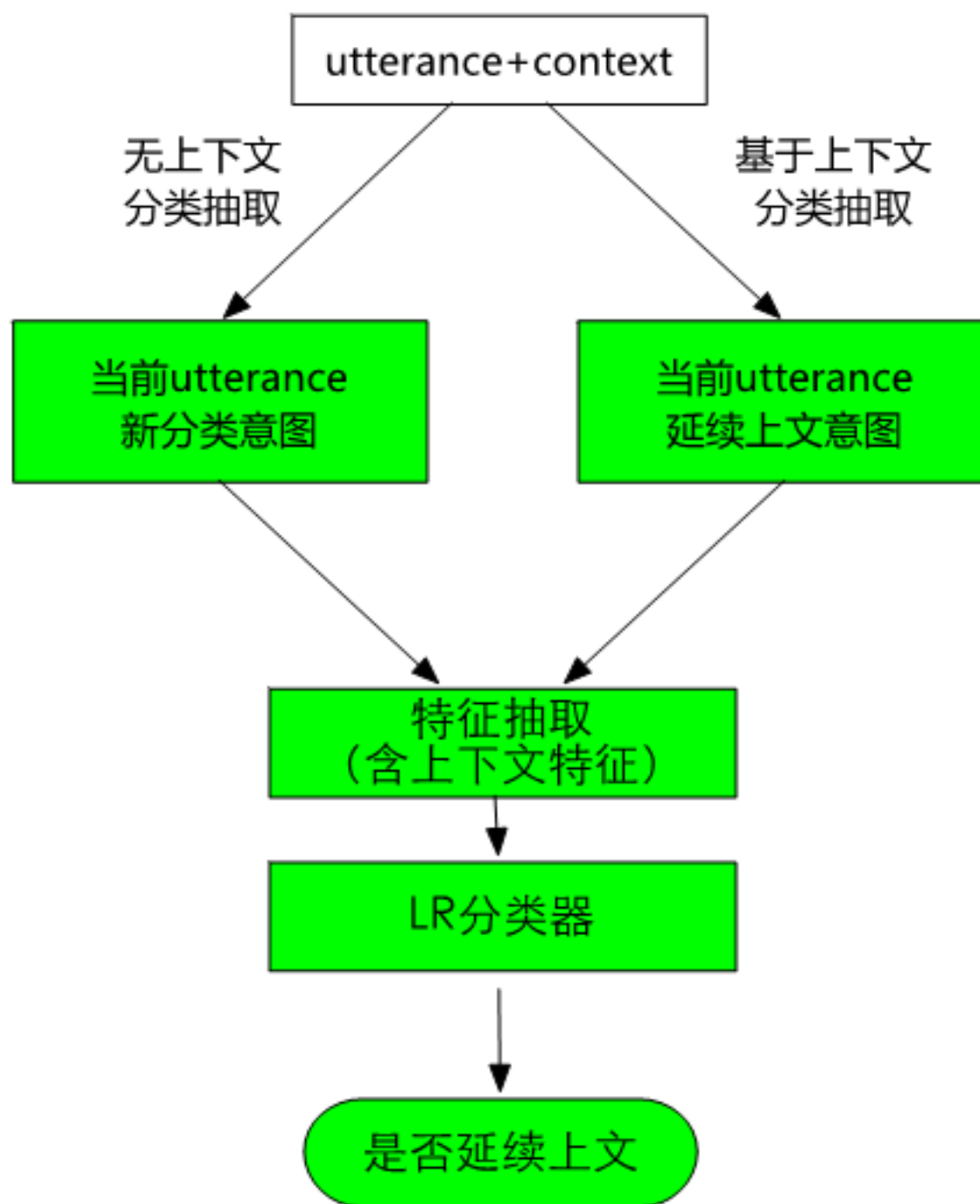


图10 基于上下文的意图延续判定

## 2.2智能问答

在具体的业务场景中有三种典型的问答任务，一是用户提供QA-Pairs，一问一答；二是建立结构化的知识图谱，进行基于知识图谱的问答；三是针对非结构化的文本，进行基于阅读理解的问答。本文重点介绍我们在阅读理解方面做的工作，比如利用阅读理解解决淘宝活动规则的问答。

在阅读理解的方法上，目前针对斯坦福大学的数据集SquAD，有大量优秀的方法不断涌现，比如match-LSTM [15]、BiDAF [16]、DCN [17]、FastQA [18]等。文献[18]给出了目前的通用框架，如图11所示，主要分为4层：① Word Embedder，对问题和文档中的词进行embedding；② Encoder，对问题和文档进行编码，一般采用RNN/LSTM/BiLSTM；③ Interaction Layer（交互层），在问题和文档之间逐词进行交互，这是目前研究的热点，主流方法是采用注意力机制（attention）；④ Answer Layer（答案层），预测答案的起始位置和结束位置。

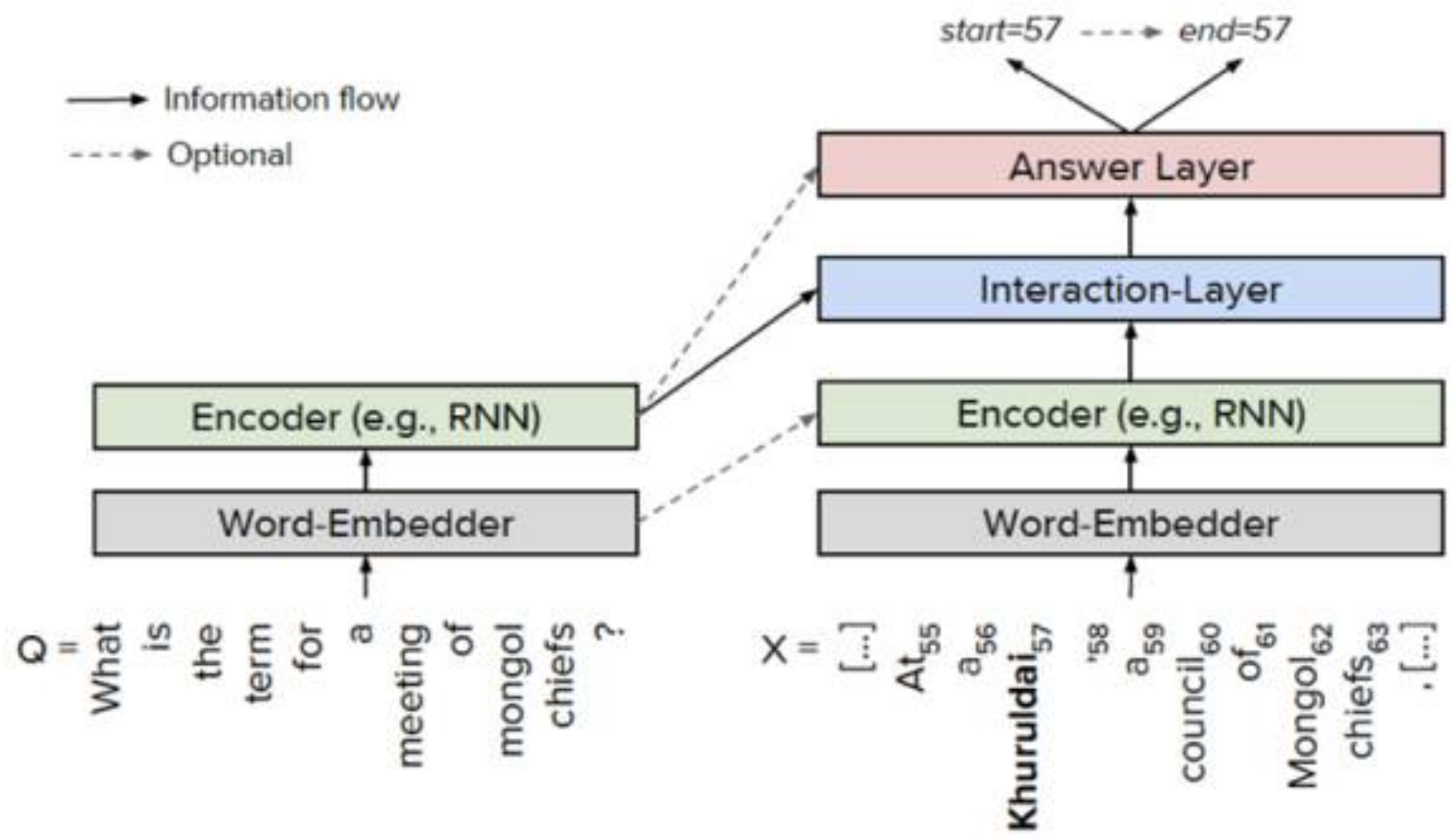


图11 阅读理解的通用框架

我们在具体实现中，参考BiDAF [16]网络结构，在此基础上做了大量优化。

### (1) 模型的业务优化

需要改进模型的结构设计，使得模型可以支持电商文档格式的输入。电商规则文档往往包含大量的文档结构，如大小标题和文档的层级结构等，将这些特定的篇章结构信息一起编码输入到网络中，将大幅提升训练的效果。



## (2) 模型的简化

学术文献中的模型一般都较为复杂，而工业界场景中由于对性能的要求，无法将这些模型直接在线上使用，需要做一些针对性的简化，使得模型效果下降可控的情况下，尽可能提升线上预测性能，例如可以简化模型中的各种bi-lstm结构。

## (3) 多种模型的融合

当前这些模型都是纯粹的end-to-end模型，其预测的可控性和可解释性较低，要适用于业务场景的话，需要考虑将深度学习模型与传统模型进行融合，达到智能程度和可控性的最佳平衡点。

## 2.3智能聊天

面向open domain的聊天机器人目前无论在学术界还是在工业界都是一大难题，目前有两种典型的方法：一是基于检索的模型，比如文献[19-20]，其基本思路是利用搜索引擎通过计算相关性来给出答案；二是基于Seq2Seq的生成式模型，典型的方法如文献[21-22]，其网络结构如图12所示。

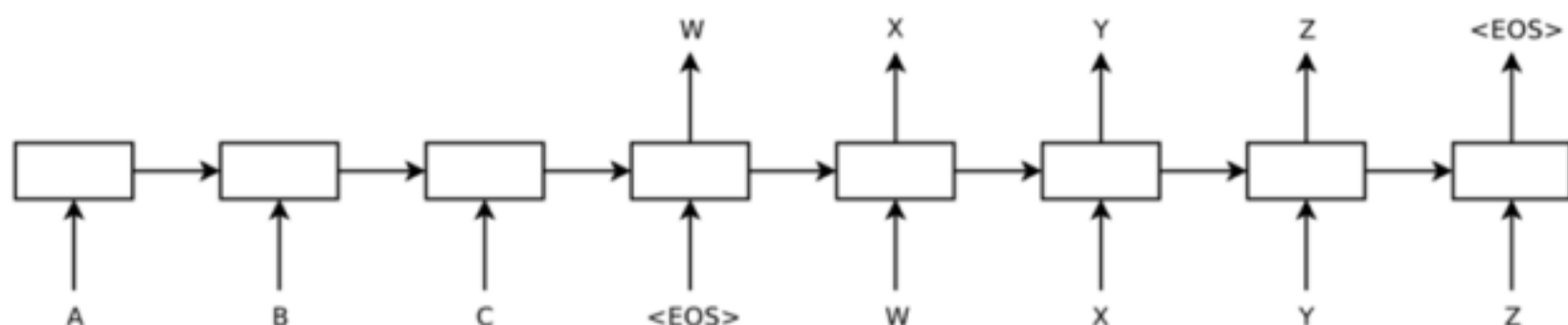


图12 Seq2Seq典型网络结构

检索模型的优点是答案在预设的语料库中，可控，匹配模型相对简单，可解释性强；缺点是在一定程度上缺乏对语义的理解，且有固定语料库的局限性，长尾问题覆盖率较差。生成模型的优点是通过深层语义方式进行答案生成，答案不受语料库规模限制；缺点是模型的可解释性不强，且难以保证回答一致性和合理性。

在我们的聊天引擎中，结合检索模型和生成模型各自的优势，提出了一种新的模型AliMe Chat [23]，基本流程如图13所示。首先采用检索模型从QA知识库中找出候选答案集合；然后利用带注意力的Seq2Seq模型对候选答案进行排序，如果第一候选的得分超过某个阈值，则作为最终答案输出，否则利用生成模型生成答案。其中带注意力的Seq2Seq模型结构如图14所示。经过训练后，主要做了如下测试：如图15所示，利用600个问题的测试集，测试了检索（IR）、生成（Generation）、检索+重排序（Rerank）及检索+重排序+生成（IR+Rerank+Generation）四种方法的效果，可以看到在阈值为0.19时，IR+Rerank+Generation的方法效果最好。

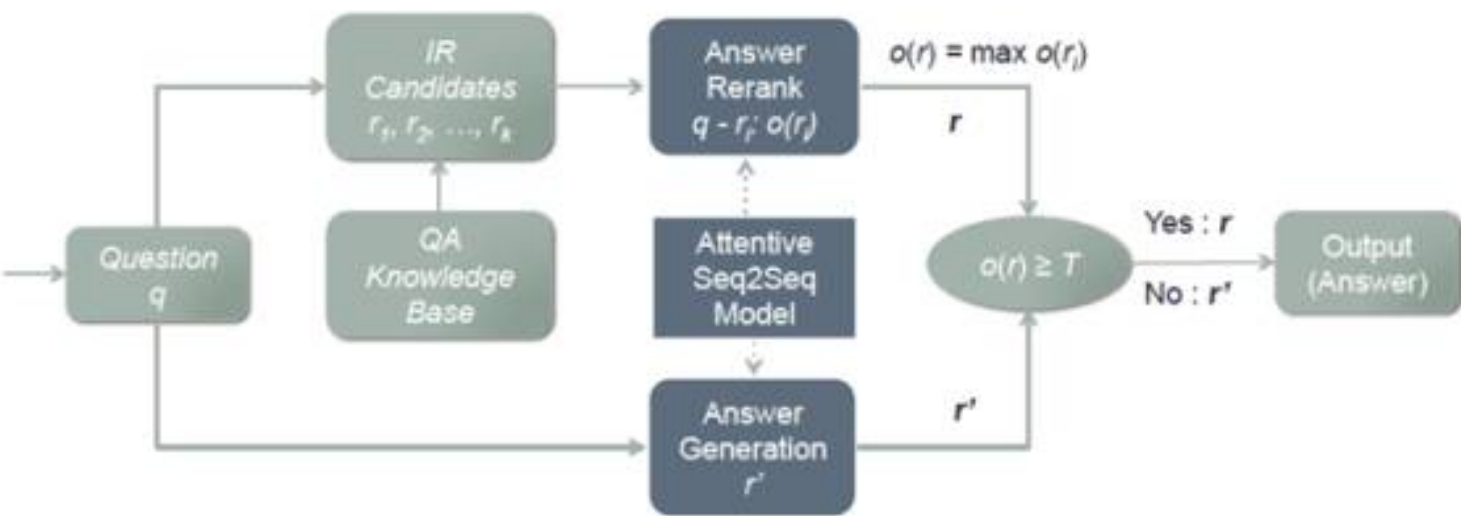


图13AliMe Chat流程图

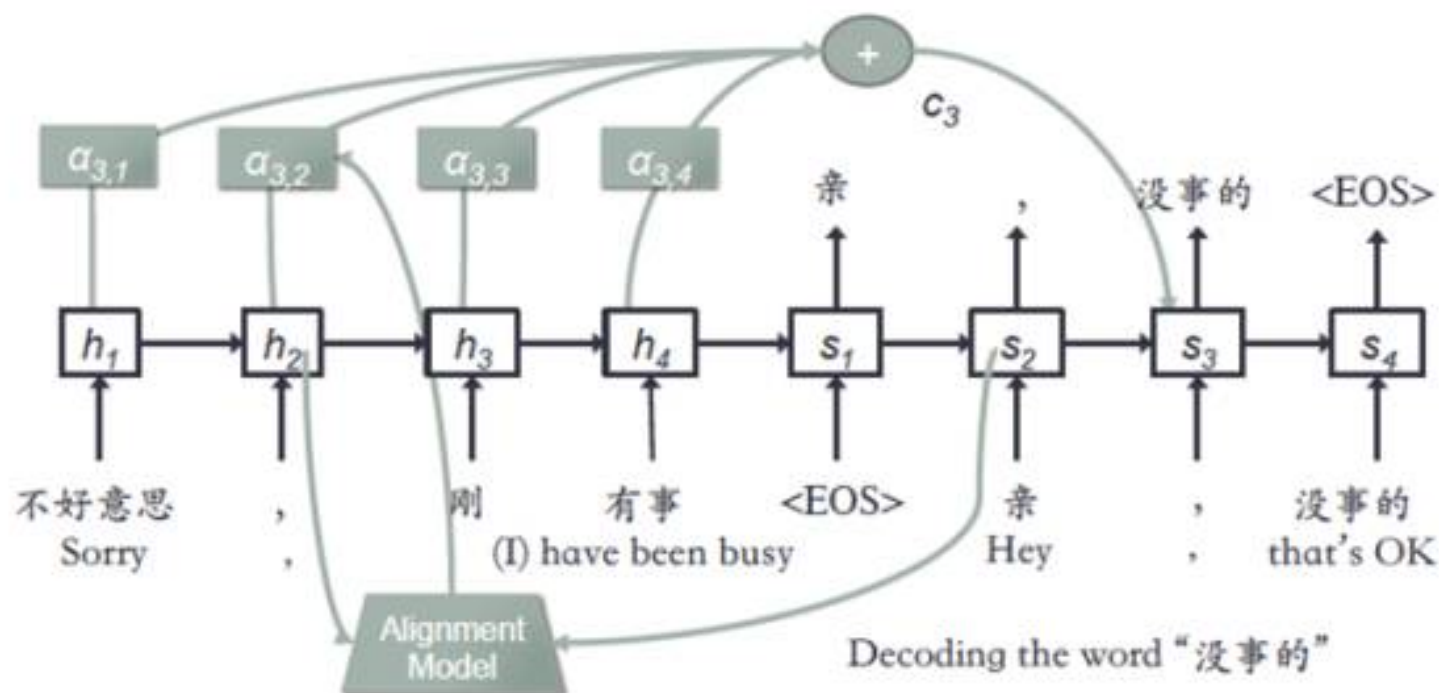


图14 带注意力的Seq2Seq网络结构示例

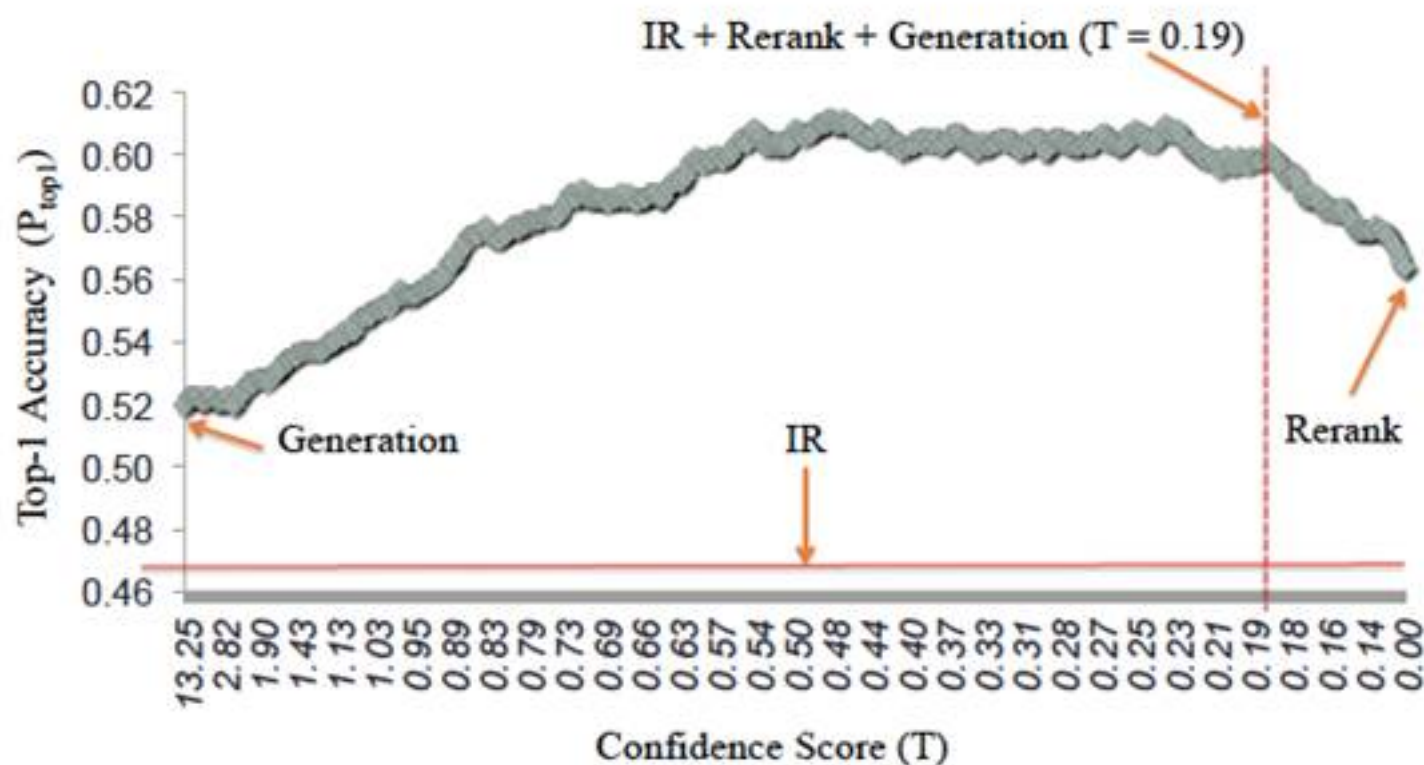


图15 IR、Generation、Rerank、IR+Rerank+Generation效果对比

此模型在阿里小蜜中上线后，示例如图16所示。在阿里小蜜中，针对之前的IR模型和AliMe Chat模型，利用线上流量做了A/B Test，结果如图17所示。从用户日志中随机选择2 136条数据，其中1 089是采用IR模型回答，另外1 047是采用AliMe Chat回答，AliMe Chat Top1答案的准确率 (accuracy) 是60.36%，远远好于IR的40.86%。

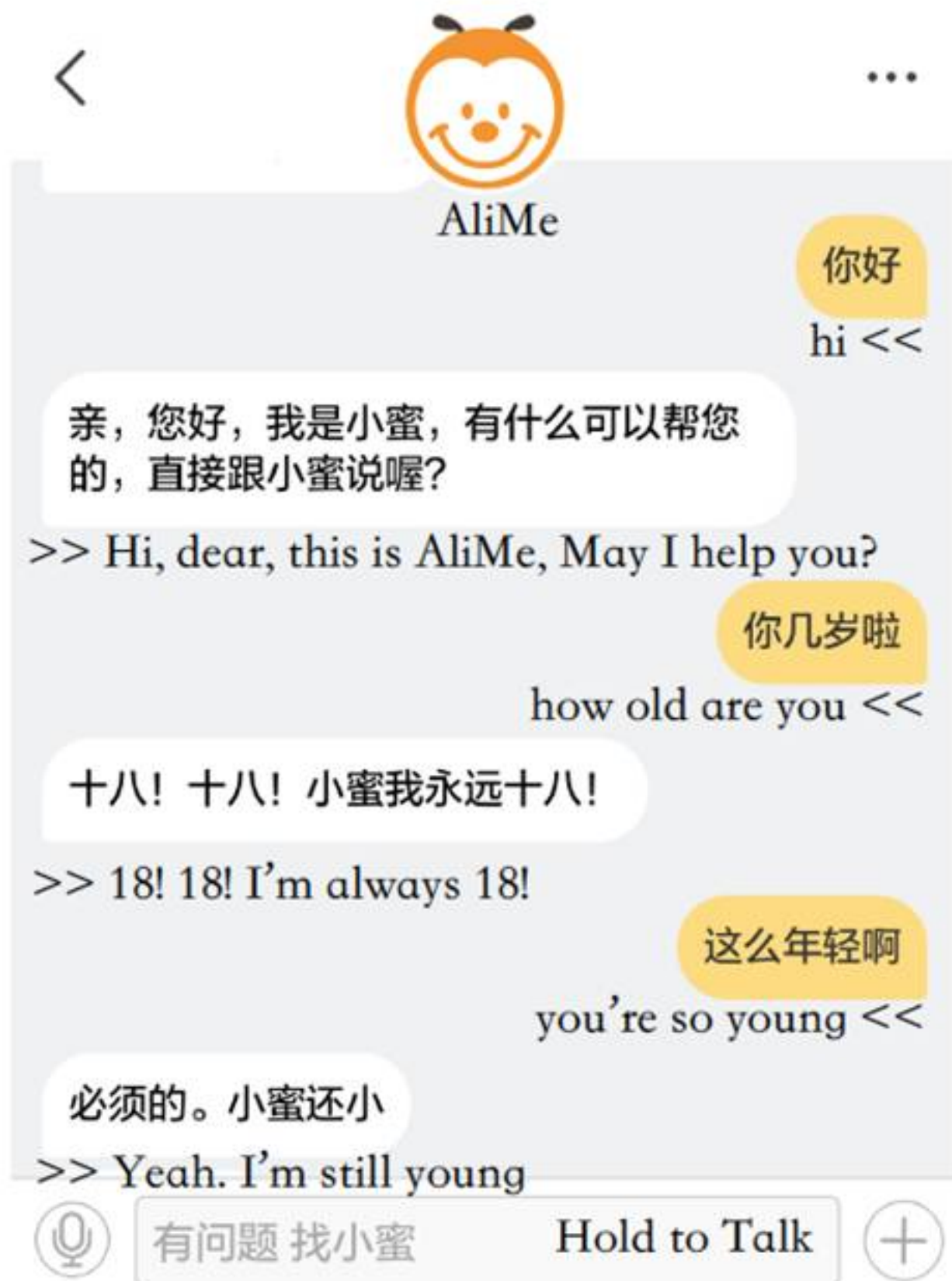


图16AliMe Chat在阿里小蜜中上线后的聊天示例

Model	$N_{total}$	$N_{unsuitable}$	$N_{neutral}$	$N_{suitable}$	$P_{top_1}$
IR	1089	644	384	61	40.86%
Hybrid	1047	415	504	128	60.36%

图17 阿里小蜜中IR方法与AliMe Chat方法A/BTest结果

## 2.4对话管理

对话管理根据语言理解的结构化语义表示结果以及上下文，来管理整个对话的状态，并决定下一步采取什么样的动作。

下面来看一个简单的对话例子。

U：我要去杭州，帮我订一张火车票

A：请问你什么时间出发？

U：明天上午

A：为你找到了以下火车票：

U：我要第二个

A：第二个是.....，您是否要购买？

U：我要购买

对话交互分成两个阶段，第一阶段，通过多轮对话交互，把用户的需求收集完整，得到结构化的信息（出发地、目的地、时间等）；第二阶段就是请求服务，接着还要去做选择、确定、支付、购买等后面一系列的步骤。

传统的人机对话，包括现在市面上常见的人机对话，一般都是只在做第一阶段的对话，第二阶段的对话做得不多。对此，我们设计了一套对话管理体系，如图18所示，这套对话管理体系具有以三个特点。



第一，设计了一套面向Task Flow的对话描述语言。该描述语言能够把整个对话任务流完整地表达出来，这个任务流就是类似于程序设计的流程图。对话描述语言带来的好处是它能够让对话引擎和业务逻辑实现分离，分离之后业务方可以开发脚本语言，不需要修改背后的引擎。

第二，由于有了Task Flow的机制，我们在对话引擎方带来的收益是能够实现对话的中断和返回机制。在人机对话当中有两类中断，一类是用户主动选择到另外一个意图，更多是由于机器没有理解用户话的意思，导致这个意图跳走了。由于我们维护了对话完整的任务流，知道当前这个对话处在一个什么状态，是在中间状态还是成功结束了，如果在中间状态，我们有机会让它回来，刚才讲过的话不需要从头讲，可以接着对话。

第三，设计了对话面向开发者的方案，称之为OpenDialog，背后有一个语言理解引擎和一个对话引擎。面向开发者的语言理解引擎是基于规则办法，能够比较好地解决冷启动的问题，开发者只需要写语言理解的Grammar，基于对话描述语言开发一个对话过程，并且还有对数据的处理操作。这样，一个基本的人机对话就可以完成了。

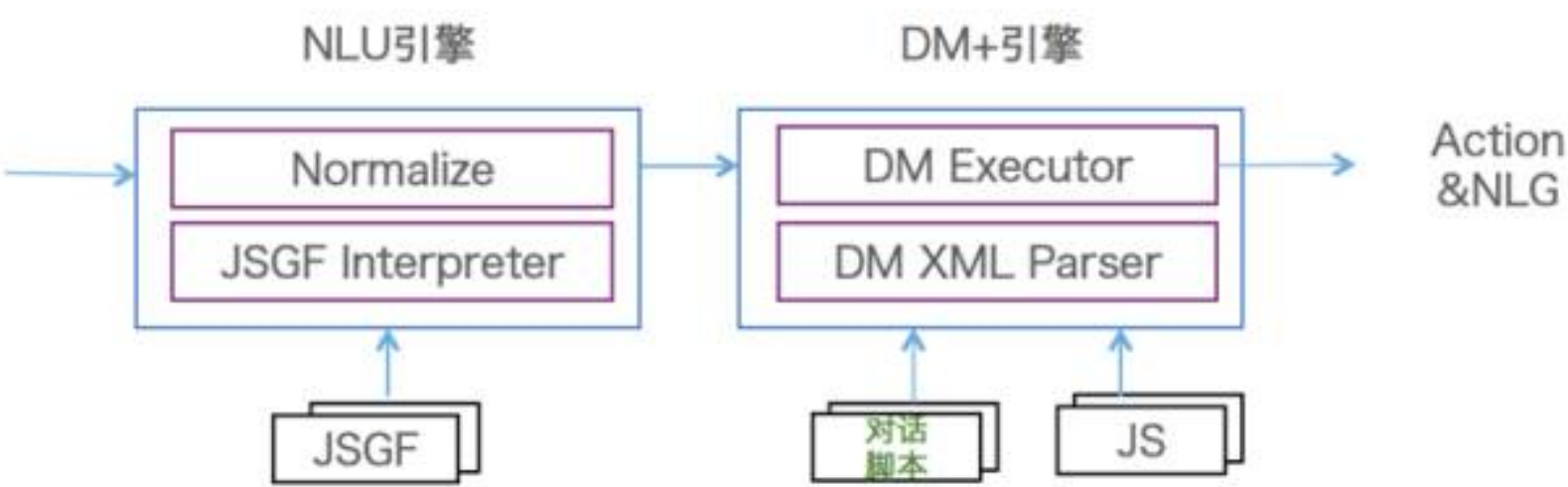


图18 对话管理框架图

### 3、阿里巴巴智能对话交互产品

### 3.1智能服务——小蜜家族

2015年7月，阿里巴巴推出了自己的智能服务助理-阿里小蜜，一个围绕着电子商务领域中的服务、导购，以及任务助理为核心的智能对话交互产品。通过电子商务领域与智能对话交互领域的结合，带来传统服务行业模式的变革与体验的提升。在2016年的双“十一”期间，阿里小蜜整体智能服务量达到643万，其中智能解决率达到95%，智能服务在整个服务量(总服务量=智能服务量+在线人工服务量+电话服务量)占比也达到95%，成为了双“十一”期间服务的绝对主力。阿里小蜜主要服务阿里国内业务和阿里国际化业务，国内业务如淘宝、天猫、飞猪、健康、闲鱼、菜鸟等，国际化业务如Lazada、PayTM、AE等。

随着阿里小蜜的成功，将智能服务能力赋能给阿里生态圈商家及阿里生态之外的企业和政府部门，便成了必然的路径。店小蜜主要赋能阿里生态中的商家，云小蜜则面向阿里之外的大中小企业、政府等。整个小蜜家族如图19所示。



图19 小蜜家族

### 3.2 智能设备

过去3~4年，我们可以看到，连接互联网的设备发生了很大变化，设备已经从PC和智能手机延伸到更广泛的智能设备，比如智能音箱、智能电视、机器人、智能汽车等设备。智能设备的快速发展正在改变着人和设备之间的交互方式。

我们研发的智能对话交互平台为各种设备提供对话交互能力，目前在YunOS手机、天猫魔盒、互联网汽车等设备上已经大量应用。比如在天猫魔盒中，用户通过对话交互可以完成搜视频、查音乐、问天气等，可以进行闲聊，还可以进行购物。

## 4 总结与思考

过去几年中，结合阿里巴巴在电商、客服、智能设备方面的刚性需求和场景，我们在智能对话交互上做了大量的探索和尝试，构建了一套相对完整的数据、算法、在线服务、离线数据闭环的技术体系，并在智能服务和智能设备上得到了大规模的应用，简单总结如下。

(1) 自然语言理解方面，通过CNN/Bi-LSTM-CRF等深度学习模型、分布式表示和符号表示的融合、多粒度的wordembedding、基于上下文的意图排序等方法，构建了规则和深度学习模型有机融合的自然语言理解系统。

(2) 智能问答方面，成功的将机器阅读理解应用在了小蜜产品中。

(3) 智能聊天方面，提出了AliMe Chat模型，融合了搜索模型和生成模型的优点，大大提高了闲聊的精度。

(4) 对话管理方面，设计了基于Task Flow的对话描述语言，将业务逻辑

辑和对话引擎分离，并能实现任务的中断返回和属性的carry-over等复杂功能。

在智能交互技术落地应用的过程，我们也在不断思考怎么进一步提高智能交互技术水平和用户体验。

第一，坚持用户体验为先。坚持用户体验为先，就是产品要为用户提供核心价值。

第二，提高语言理解的鲁棒性和领域扩展性。

第三，大力发展机器阅读理解能力。

第四，打造让机器持续学习能力。

第五，打造数据闭环，用数据驱动效果的持续提升。

目前的人工智能领域仍然处在弱人工智能阶段，特别是从感知到认知领域需要提升的空间还非常大。智能对话交互在专有领域已经可以与实际场景紧密结合并产生巨大价值，尤其在智能客服领域（如阿里巴巴的小蜜）。随着人工智能技术的不断发展，未来智能对话交互领域的发展还将会有不断的提升。

阿里巴巴集团-智能服务事业部招募自然语言理解、人机对话、知识图谱和智能问答等方向的人才啦！该职位将致力于自然语言的基础研究和开发（NLP）、自然语言的语义理解，打造全球领先的智能人机对话交互平台（阿里小蜜、店小蜜、云小蜜），服务于各行各业的企业/组织。研发方向包括但不限于：

1. 自然语言理解算法和技术的研究；
2. 人机对话模型的算法和技术的研究；

3. 知识图谱和智能问答相关算法和技术的研究；

4. 针对人工智能客服机器人的应用和产品研发；

如果你愿意和我们一起创造人机交互的未来，热烈欢迎加入我们！

联系方式：jian.sun@alibaba-inc.com(千诀)

## 参考文献

**[1] [https://en.wikipedia.org/wiki/Natural\\_language\\_understanding](https://en.wikipedia.org/wiki/Natural_language_understanding).**

**[2]Percy Liang, Natural Language Understanding:Foundations and State-of-the-Art,ICML, 2015.**

[3]Yoon Kim, Neural Networks for Sentence Classification, EMNLP, 2014.

[4] Suman Ravuri, and Andreas Stolcke, Recurrent Neural Network and LSTM Modelsfor Lexical Utterance Classification, InterSpecch, 2015.

[5] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao, Recurrent ConvolutionalNeural Networks for Text Classification, AAAI, 2015.

[6] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis C.M. Lau, A C-LSTMNeural Network for Text Classification, arXiv, 2015.

[7]Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov, Bag ofTricks for Efficient Text Classification, EACL, 2017.

[8] Piotr Bojanowski,Edouard Grave, Armand Joulin and Tomas Mikolov, EnrichingWord Vectors withSubword Information, TACL, 2017.

[9] C. Raymond, and G.Riccardi, Generative and discriminative algorithms for spoken languageunderstanding, Interspeech, 2007.



- [10] Kaisheng Yao, Geoffrey Zweig, Mei-Yuh Hwang, Yangyang Shi, and Dong Yu, Recurrent neural networks for language understanding, InterSpeech, 2013.
- [11] Kaisheng Yao, Baolin Peng, Geoffrey Zweig, Dong Yu, Xiaolong Li, and Feng Gao, Recurrent conditional random field for language understanding, ICASSP, 2014.
- [12] Kaisheng Yao, Baolin Peng, Yu Zhang, Dong Yu, Geoffrey Zweig, and Yangyang Shi, Spoken language understanding using long short-term memory neural networks, 2014 IEEE Spoken Language Technology Workshop (SLT), 2014.
- [13] Grégoire Mesnil, Yann Dauphin, Kaisheng Yao, Yoshua Bengio, Li Deng, Dilek Hakkani-Tur, Xiaodong He, Larry Heck, Gokhan Tur, Dong Yu, and Geoffrey Zweig, Using Recurrent Neural Networks for Slot Filling in Spoken Language Understanding, TASLP, 2015
- [14] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer, Neural architectures for named entity recognition, NAACL, 2016.
- [15] Shuohang Wang, and Jing Jiang, Machine comprehension using match-lstm and answer pointer, ICLR, 2017.
- [16] Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hananneh Hajishirzi, Bidirectional attention flow for machine comprehension, ICLR, 2017.
- [17] Caiming Xiong, Victor Zhong, and Richard Socher, Dynamic coattention networks for question answering, ICLR, 2017.
- [18] Dirk Weissenborn, Georg Wiese, and Laura Seiffe, Making Neural QA as Simple as Possible but not Simpler, arXiv, 2017.
- [19] Zongcheng Ji, Zhengdong Lu, and Hang Li, An information retrieval approach to short text conversation, arXiv, 2014.

[20] Zhao Yan, NanDuan, Jun-Wei Bao, Peng Chen, Ming Zhou, Zhoujun Li, and Jianshe Zhou, Docchat:An information retrieval approach for chatbot engines using unstructureddocuments, ACL, 2016.

[21] Dzmitry Bahdanau,Kyunghyun Cho, and Yoshua Bengio, Neural machine translation by jointly learningto align and translate, ICLR, 2015.

[22] Oriol Vinyals, andQuoc V. Le., A neural conversational model, ICML Deep Learning Workshop, 2015.

[23] Minghui Qiu,Feng-Lin Li, Siyu Wang, et al., AliMe Chat: A Sequence to Sequence and Rerankbased Chatbot Engine, ACL, 2017.

----- End -----

你可能还喜欢

[多隆：从工程师到合伙人](#)

[阿里技术人纪录片](#)



[如何降低90%Java垃圾回收时间？](#)

[以阿里HBase的GC优化实践为例](#)



[燃！阿里AI技术取得重大突破：](#)

[连破中、英语言处理两项世界纪录](#)



关注「阿里技术」

把握前沿技术脉搏