多轮对话下的多标签分类

摘要

Spoken language understanding (SLU)是对话系统的核心组成部分，其中包括领域分类，意图分类和槽填空[1]。而意图分类属于多标签文本分类问题，是目前对话系统的难点问题。传统的单轮对话下的意图分类存在两个较大的问题，第一点：一个意图是很难通过一句话完成，这使得单轮的意图分类变得困难。第二点：一个意图可能和之前的意图有关系，这使得单独的一轮对话很难进行准确的意图识别。我们将提出一种模型来处理上面的两个问题，基于多轮的对话情景来完成意图分类，识别对话的多个意图。

介绍

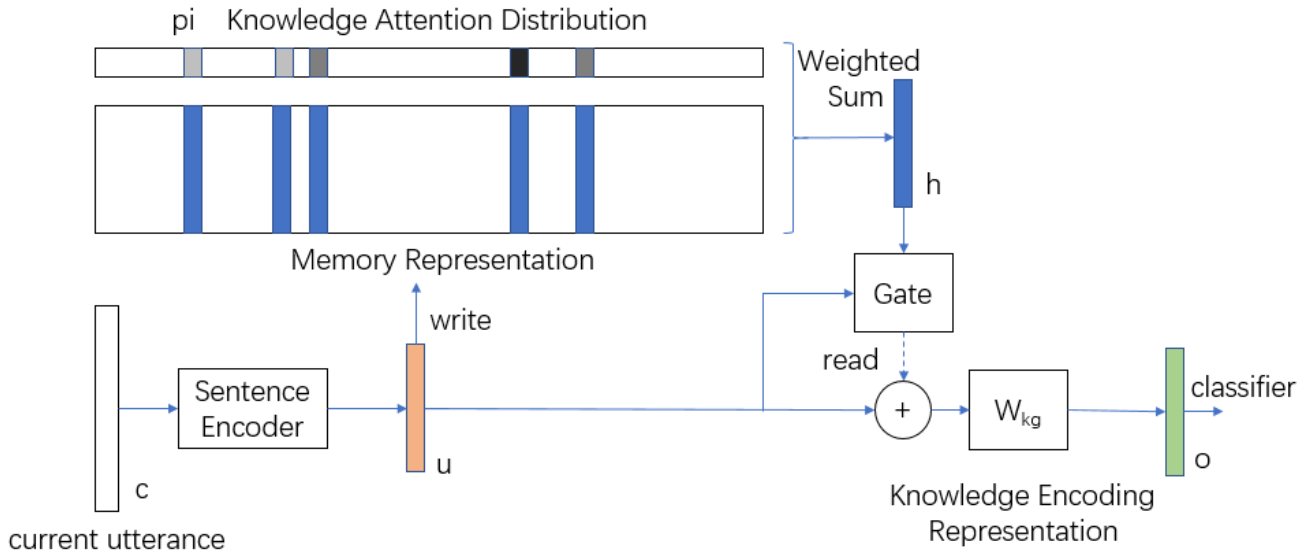
对话系统是人工智能和自然语言处理的研究人员长久以来倍感兴趣研究但是应用效果不佳的一个应用研究方向。对话系统可以用来解放大量的客户服务人员，如帮助预定电影票，解决售后预约，充当导游等。多轮对话下的意图识别是多标签文本分类的一个应用，目前是热门研究的方向。微软的小冰，百度的度秘，苹果的Siri等聊天机器人，以及一些任务驱动的多轮对话机器人都把Spoken language understanding (SLU)中的意图分类作为最基础最重点的研究之一。所谓的意图包括说话者的语言行为和相关属性。意图的单轮对话的意图分类定义为：给定当前对话内容ui，目标是预测出ui的意图，如：u1：明天天气怎么样？(意图：询问_天气)。多轮对话的意图分类定义为：给定当前对话ui和之前的聊天内容vi，目标是预测出ui的意图，如：u1：明天天气怎么样？(意图：询问_天气)；u2：后天呢？(意图：询问_天气)。这里如果没有上文的信息，u2的意图分类是比较模糊的。

传统的文本分类方法有支持向量机（SVM）和最大熵等，过去常被用来做新闻等长文本分类，其应用到对话系统上，在对话片段的领域分类上，依旧有不错的表现[2][3][4]。但应用于对话系统的意图分类这类短文本分类上，效果较差。随着深度学习的发展，深度神经网络（DNNs）被应用于意图分类[5][6][7]。最近Ravuri et al.提出一个RNN框架来解决意图分类问题[8]；Xu et al.提出利用CNN框架结合CRF模型同时解决意图分类和槽填空问题[9]；Hakkani-Tur et al.提出一个基于RNN框架的多个领域多个任务联合学习的模型[10]。然而这些模型都聚焦于单轮的对话系统，每个对话的意图都是相互独立的。这导致模型在很多应用场景中受到限制，不能被广泛应用于现实场景中。

上下文的信息对于理解当前对话的意图有着重要的作用，如上文提到的多轮对话的例子，u2的意图被预测为询问_天气需要依靠u1对话中的信息，这种省略在口语对话中非常常见。在这种情况下，需要利用上文对话提供必要的信息才能判断当前意图。 Hakkani-Tur et al.提出SVM-HMMs从上文的对话中抽取信息联合当前对话来实现槽填空和意图识别[11]；Xu et al.提出利用RNNs框架结合上文的对话信息来实现领域分类和意图识别[12]。然而过去的这些方法仅仅利用上文对话内容抽取信息，忽略了长依赖和主次信息，这将导致上文对话信息的利用率降低。

最近，有一些计算模型会使用存储器和注意力概念来提升模型效果[13][14][15]。存储器可以允许众多的复杂的计算步骤和长依赖在序列对话中。基本上，存储器结合神经网络构成连续的表示模型，将编码的知识信息通过读和写的操作更新存储器中的信息。注意力机制可以将从存储器中读取的信息分为主要信息和次要信息，从而使得历史信息的利用率提高，提高模型效果。Hakkani-Tur et al.提出一个END-To-END的RNNs框架利用存储器和注意力机制实现多轮对话的意图分类[15]。然后模型解决了当前意图和上文相关的问题，但是对于出现的新的意图，与上文不相关的情况下，模型会因为引入的历史信息而引入大量的噪声。我们提出一个新的改进模型，在模型读取存储器的时候，为模型添加一个门，来控制在判断当前对话意图的时候，是否需要引入历史信息。如果意图与历史信息相关，门通过读取历史信息，来帮助当前对话意图的识别。如果意图与历史信息不相关，门拒绝读取历史信息，仅仅依靠当前对话识别意图。这样做，可以综合考虑多个意图的出现的多种情况，防止在某些情况下引入噪声。

模型



参考文献

[1] Gokhan Tur and Renato De Mori. 2011. Spoken language understanding: Systems for extracting semantic information from speech. John Wiley & Sons.

[2] P. Haffner, G. Tur, and J. H. Wright, “Optimizing svms for complex call classification,” in 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP), vol. 1. IEEE, 2003, pp. I–632.

[3] C. Chelba, M. Mahajan, and A. Acero, “Speech utterance classification,” in 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP), vol. 1. IEEE, 2003, pp. I–280.

[4] Y.-N. Chen, D. Hakkani-Tur, and G. Tur, “Deriving local relational surface forms from dependency-based entity embeddings for unsupervised spoken language understanding,” in 2014 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2014, pp. 242–247.

[5] R. Sarikaya, G. E. Hinton, and B. Ramabhadran, “Deep belief nets for natural language call-routing,” in 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011, pp. 5680–5683.

[6] G. Tur, L. Deng, D. Hakkani-Tur, and X. He, “Towards deeper un- ¨ derstanding: Deep convex networks for semantic utterance classification,” in 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2012, pp. 5045–5048.

[7] R. Sarikaya, G. E. Hinton, and A. Deoras, “Application of deep belief networks for natural language understanding,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 4, pp. 778–784, 2014.

[8] S. Ravuri and A. Stolcke, “Recurrent neural network and lstm models for lexical utterance classification,” in Sixteenth Annual Conference of the International Speech Communication Association, 2015.

[9] P. Xu and R. Sarikaya, “Convolutional neural network based triangular CRF for joint intent detection and slot filling,” in 2013 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). IEEE, 2013, pp. 78–83.

[10] Yun-Nung Chen, Dilek Hakkani-Tur, Gokhan Tur, Jianfeng Gao, Li Deng and Ye-Yi Wang, “Multi-domain joint semantic frame parsing using bi-directional RNN-LSTM,” In Proceedings of The 17th Annual Meeting of the International Speech Communication Association. pages 715–719.

[11] A. Bhargava, A. Celikyilmaz, D. Hakkani-Tur, and R. Sarikaya, “Easy contextual intent prediction and slot detection,” in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2013, pp. 8337–8341.

[12] P. Xu and R. Sarikaya, “Contextual domain classification in spoken language understanding systems using recurrent neural network,” in 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014, pp. 136–140.

[13] J. Weston, S. Chopra, and A. Bordesa, “Memory networks,” in International Conference on Learning Representations (ICLR), 2015.

[14] S. Sukhbaatar, J. Weston, R. Fergus et al., “End-to-end memory networks,” in Advances in Neural Information Processing Systems, 2015, pp. 2431–2439.

[15] Yun-Nung Chen, Dilek Hakkani-Tur, Gokhan Tur, Jianfeng Gao, and Li Deng. 2016c. End-to-end memory networks with knowledge carryover for multi-turn spoken language understanding. In Proceedings of The 17th Annual Meeting of the International Speech Communication Association. pages 3245–3249.