
软件所在开源社区群智挖掘研究中获进展

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/16223.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

近日，中国科学院软件研究所互联网软件技术实验室在开源社区群智挖掘研究中取得进展。该研究面向开源社区中的开发者，针对开发者实时聊天内容中大量有价值的“问题-解决方案”信息，提出了一种多层次模型——面向群智的问题及解决方案自动提取技术（ISPY），有效构建了“问题-解决方案”（Issue-Solution）知识库，实现了当前“问题-解决方案”信息提取性能最优，在促进开源社区知识共享、提升问题解决效率等方面具有重要意义。

Gitter、Slack等开源社区实时聊天平台是目前开发者协作的主要沟通工具，在软件开发和维护的过程中被广泛使用。社区聊天中包含了大量有价值的问题-解决方案信息，这些信息能够有效地提高软件质量和生产力。比如，开发者可以将项目开发过程中存在的安装、编译等问题推送到平台上，经由其他开发者的回复以解决这些问题。而由于社区群体聊天的对话存在耦合性和复杂性，现有方法难以对其挖掘和提取。

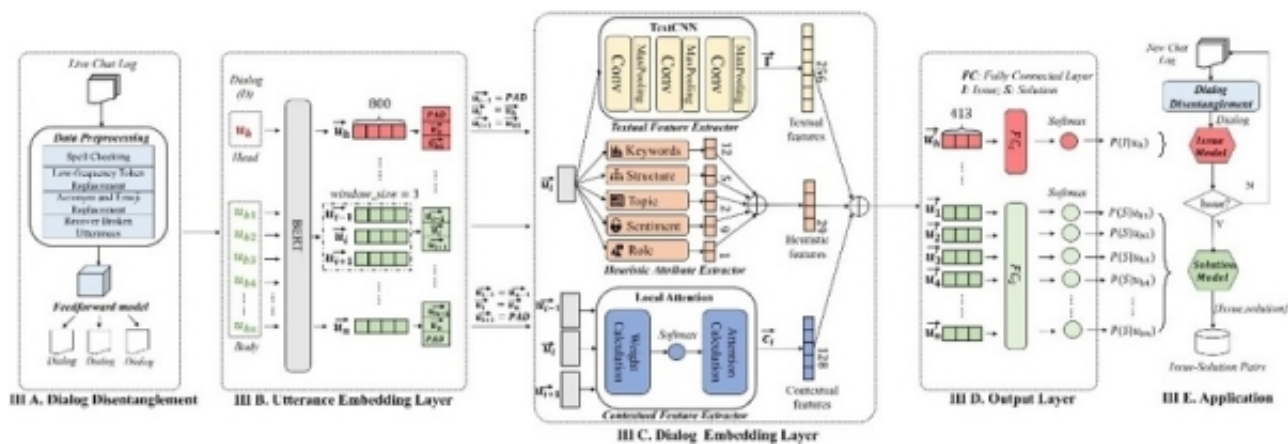
科研团队提出的ISPY模型利用多层前馈网络模型将耦合的对话进行解耦，形成若干独立的对话；模型基于卷积网络提取对话的文本特征，基于启发式规则提取语义特征，基于局部注意力机制提取对话的上下文特征。在预测阶段，进行两步预测，第一步预测对话内容是否讨论问题（问题识别），第二步预测对话的内容哪些属于解决方案（解决方案抽取）。

研究团队在8个开源社区的4944条聊天数据上测试ISPY模型的准确率、召回率以及调和平均值。在问题识别任务上，ISPY模型超过所有现有指标，调和平均值达到76%，平均提升当前最优基线30%；在解决方案抽取任务上，准确率与召回率指标超过现有基线水平，调和平均值达到63%，平均提升最优基线20%。此外，研究团队还利用ISPY模型实现了在StackOverflow平台上自动问答，以6/26的最佳答案、19/26的候选答案被该平台采纳，进一步证明了该模型的有效性和实用性。

相关研究成果以*ISPY: Automatic Issue-Solution Pair Extraction from Community Live Chats*为题，被软件工程领域会议ASE2021收录，并获ACM SIGSOFT杰出论文奖。研究工作得到国家自然科学基金青年科学基金、中科院青年创新促进会、软件所优秀青年科技人才计划的支持。

[论文链接](#)

[代码链接](#)



面向群智的问题及解决方案自动提取方法框架

研究团队单位：软件研究所

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发