

Detection of Hidden Feature Requests from Massive Chat Messages via Deep Siamese Network

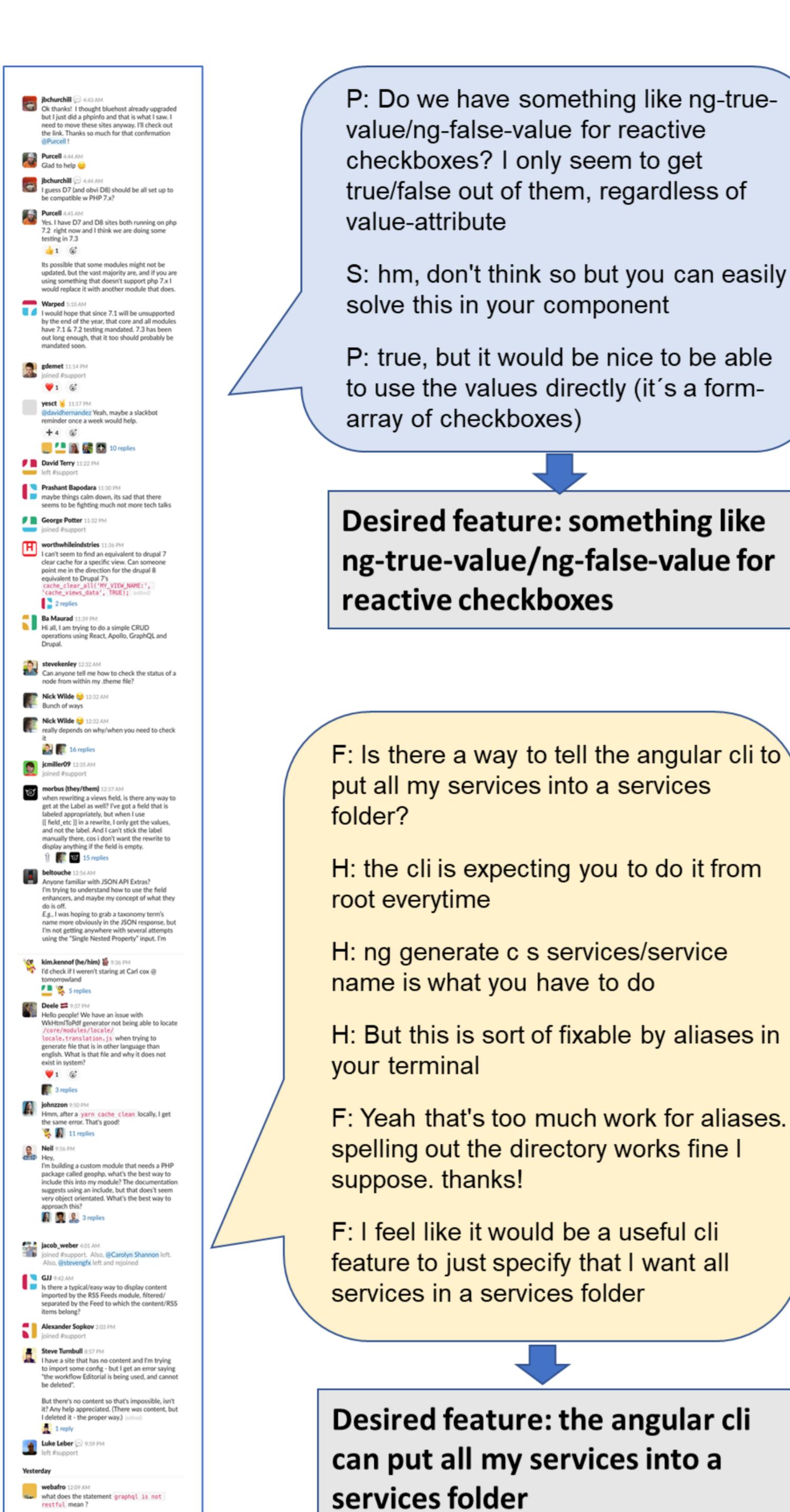
作者：石琳 邢铭哲 李明阳 王亚文 李守斌 王青

Research Paper in the International Conference on Software Engineering (ICSE2020)

主要联系人：石琳 (15001193593, shilin@iscas.ac.cn)

Introduction

- 在开源项目和工业软件中，开发者会大量使用共享交流平台，如邮件列表、issue tracker、在线聊天工具等。其中在线聊天逐渐成为重要的开发者讨论平台，开发人员经常会在对话中向其他开发者提及所期望的功能，因此聊天记录中存在着大量关于需求信息的对话。然而从大规模聊天消息中自动检索需求信息存在多种困难，因为对话通常长度较长且掺杂无关信息，较难对聊天内容有透彻的了解，并且从大量的聊天记录中识别并标注出需求信息的对话如大海捞针一般，是一项时间成本、人力成本昂贵的任务。
- 为此，我们提出了FRMiner，该方法以我们设计的对话模型为基模型构建孪生网络，其核心思想在于将传统的文本分类任务转换为确定两个对话是否是相似的任务，通过少样本学习技术建立分类模型。
- FRMiner取得了88.52%、88.50%和88.51%的平均精度、召回率和F1值，超越了以往需求挖掘方法和文本分类模型。该方法可以帮助收集全面的需求进而减少开发者大量重复工作。另外，我们开源了标注的1,035条数据以及代码，可供复现和未来进一步研究。

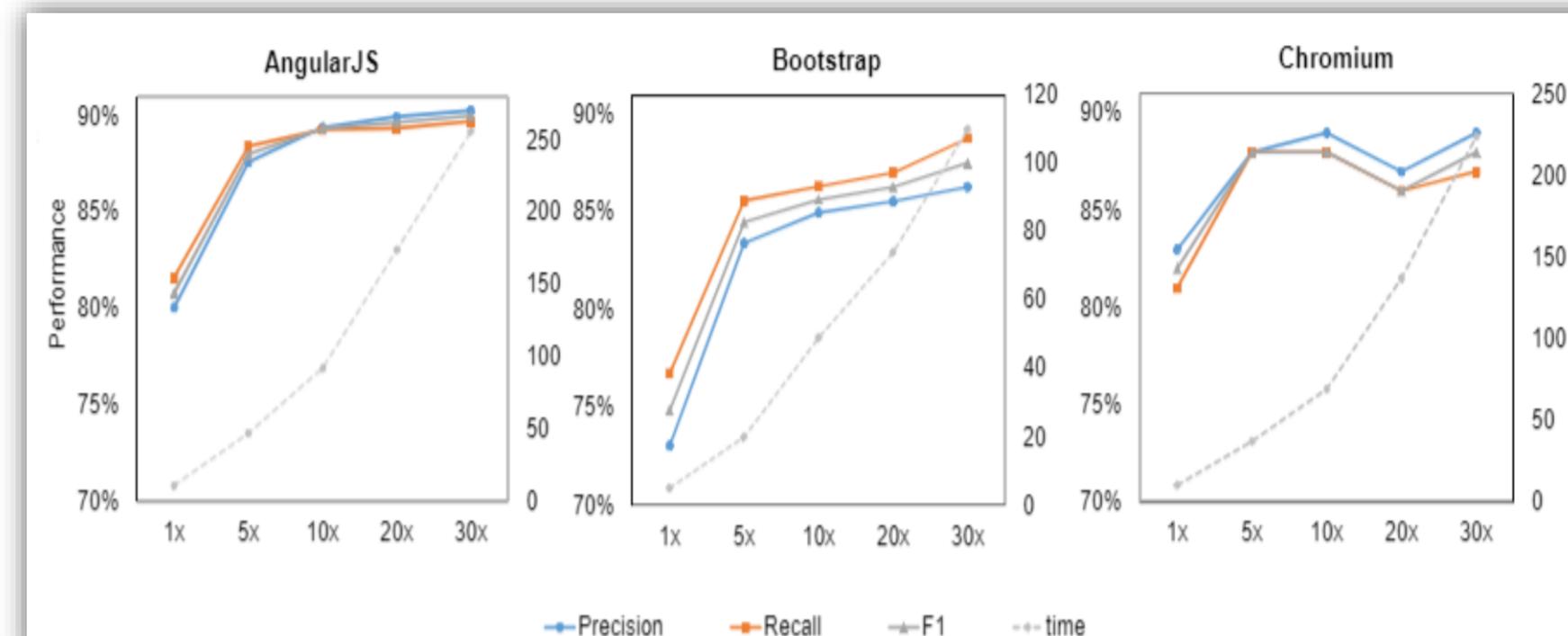


AngularJS项目中的一个实际例子，软件需求隐藏在对话过程中

Experiments

方法	AngularJS			Bootstrap			Chromium		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
本文方法	FRMiner	90.28%	89.73%	90.00%	86.28%	88.78%	87.52%	89.00%	87.00%
	p-FRMiner	31.71%	54.17%	40.00%	50.00%	47.80%	48.98%	14.00%	44.00%
已有研究方法	CNC	7.70%	44.44%	13.13%	16.38%	34.21%	22.13%	9.56%	67.00%
	FRA	13.67%	80.33%	23.35%	23.00%	48.67%	31.00%	12.00%	20.00%
文本分类方法	NB	20.00%	27.67%	22.33%	25.67%	62.00%	36.00%	14.33%	44.33%
	GBDT	36.00%	22.33%	27.33%	41.67%	35.67%	38.33%	9.33%	7.33%
	RF	52.67%	11.00%	16.33%	57.00%	29.00%	38.33%	0.00%	0.00%
	FT	23.33%	5.33%	8.67%	57.67%	29.00%	38.33%	38.00%	9.10%

项目内三折交叉验证，并与基线方法对比



FRMiner和p-FRMiner对比结果

方法	AngularJS			Bootstrap			Chromium		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
本文方法	FRMiner	85.23%	86.56%	85.89%	86.84%	85.89%	86.37%	85.87%	86.81%
	p-FRMiner	31.03%	50.00%	38.30%	27.56%	69.08%	39.40%	16.00%	24.24%
已有研究方法	CNC	7.70%	44.44%	13.13%	16.38%	34.21%	22.13%	9.56%	67.00%
	FRA	13.67%	80.33%	23.35%	23.00%	48.67%	31.00%	12.00%	20.00%
文本分类方法	NB	16.00%	75.00%	26.00%	27.00%	36.00%	31.00%	7.00%	26.00%
	GBDT	18.00%	14.00%	16.00%	30.00%	11.00%	16.00%	20.00%	19.00%
	RF	28.00%	14.00%	19.00%	37.00%	9.00%	15.00%	12.00%	26.00%
	FT	32.00%	19.00%	24.00%	43.00%	13.00%	20.00%	19.00%	14.00%

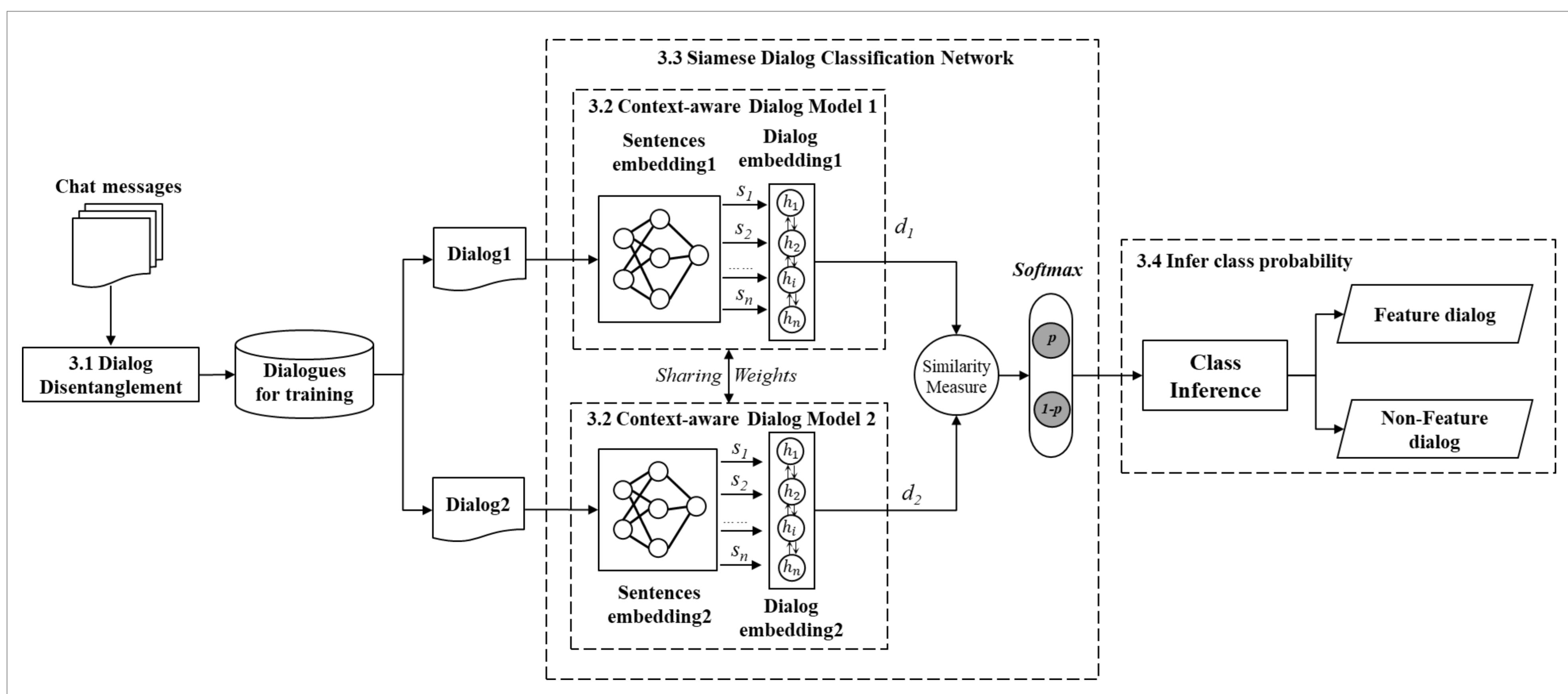
跨项目泛化性实验结果

与基线对比实验表明FRMiner效果明显高于近年先进的一些研究方法。其原因在于模型判断两个对话类别是否相似要比对单个对话进行分类更加容易，并且数据增强保证了训练数据的充足性。

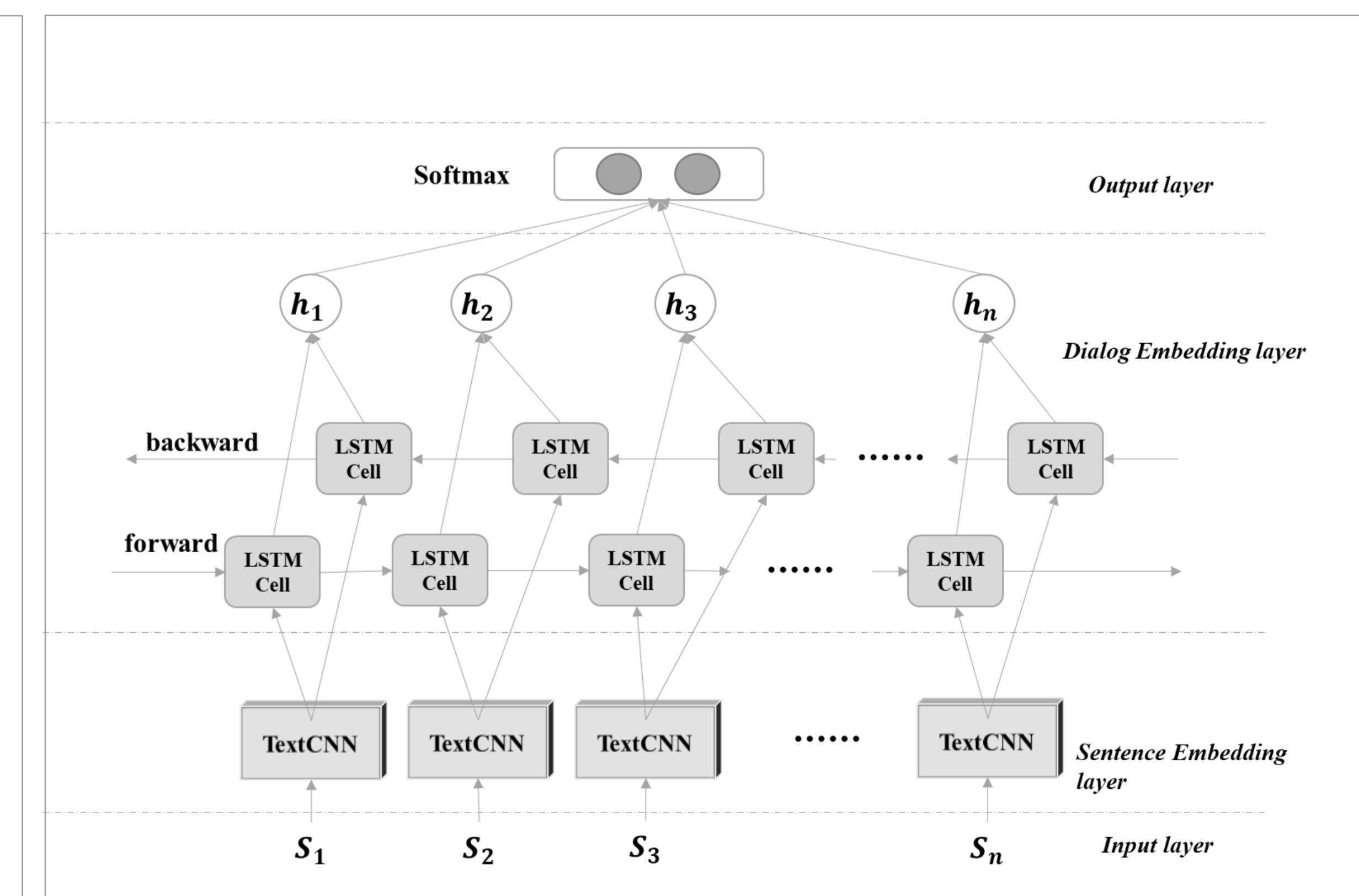
FRMiner和无孪生网络的p-FRMiner实验对比证明我们方法的高效。另外5倍的数据增强是模型效果和训练成本之间较好的trade-off值。

跨项目实验证明开发者在不同的社区和项目中倾向于使用相似的模式表达软件的需求，并且FRMiner在提取需求表达范式并将其泛化到其他领域有较好的性能。

Model



- 以对话模型为基模型构建包含相似度度量层的孪生网络
- 构建Pair-Instance，数据从N增强到O(N²)级别
- 根据模型预测结果和配对对话的标签进行目标对话类别推断



分层的上下文敏感对话模型。其以TextCNN和BiLSTM分别对句子和对话进行建模

Conclusion

我们提出了一种使用孪生网络从大规模聊天信息中识别软件需求的方法FRMiner。其结合两个相同结构、共享参数的孪生上下文敏感对话模型去学习对话之间的相似性，而不是传统的对单个对话进行分类的方法。在三个开源项目上的实验结果表明FRMiner的效果较大程度上超越了两个句子级别的需求识别方法和四个文本分类模型；跨项目实验中证明了FRMiner也具有较好的泛化能力，能将学习到的需求表达模式泛化到其他项目领域。