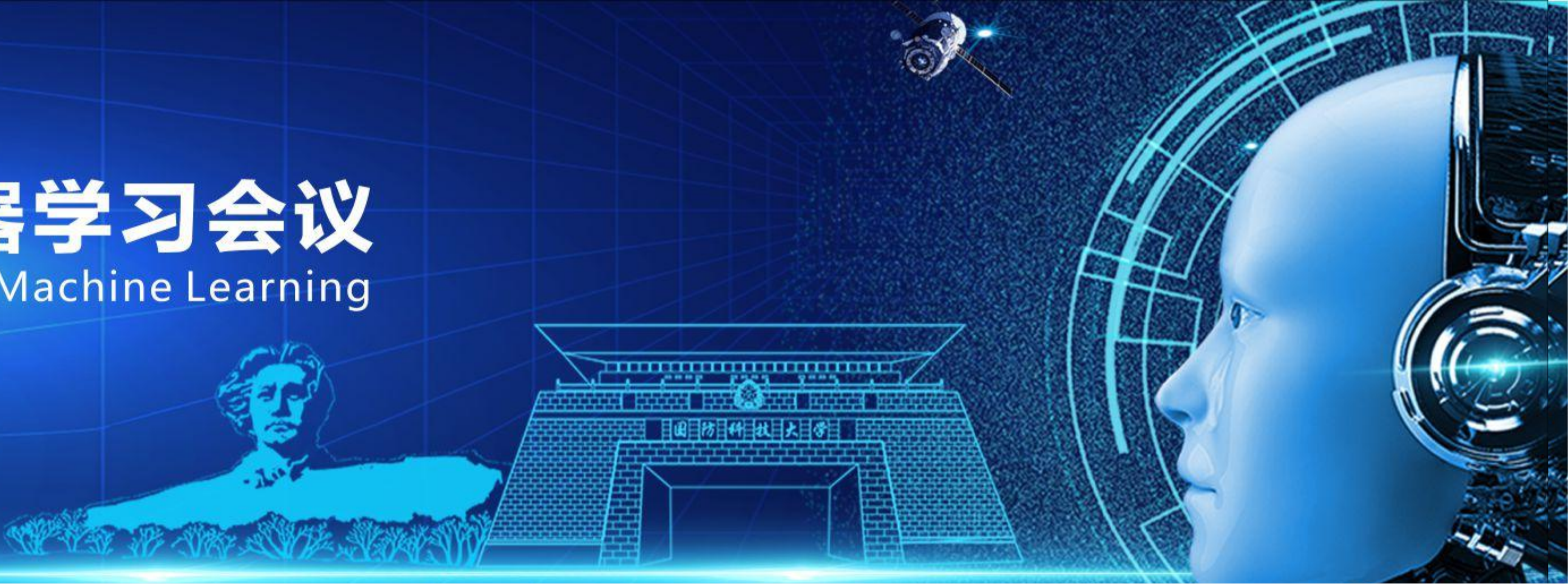


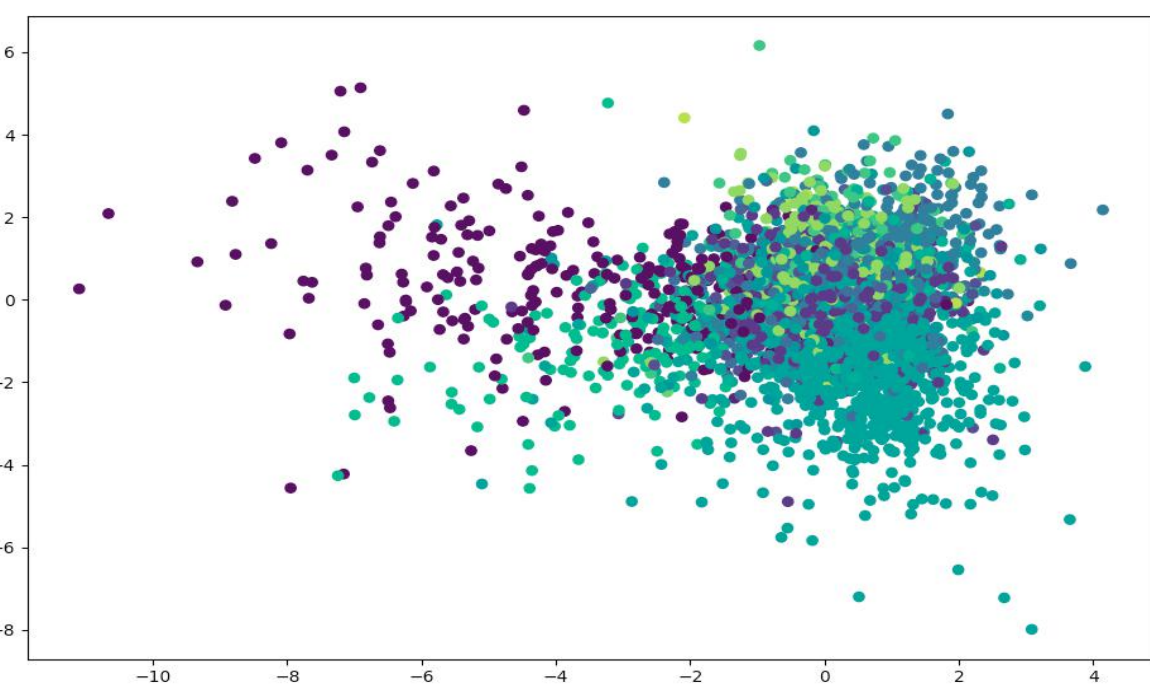


摘要

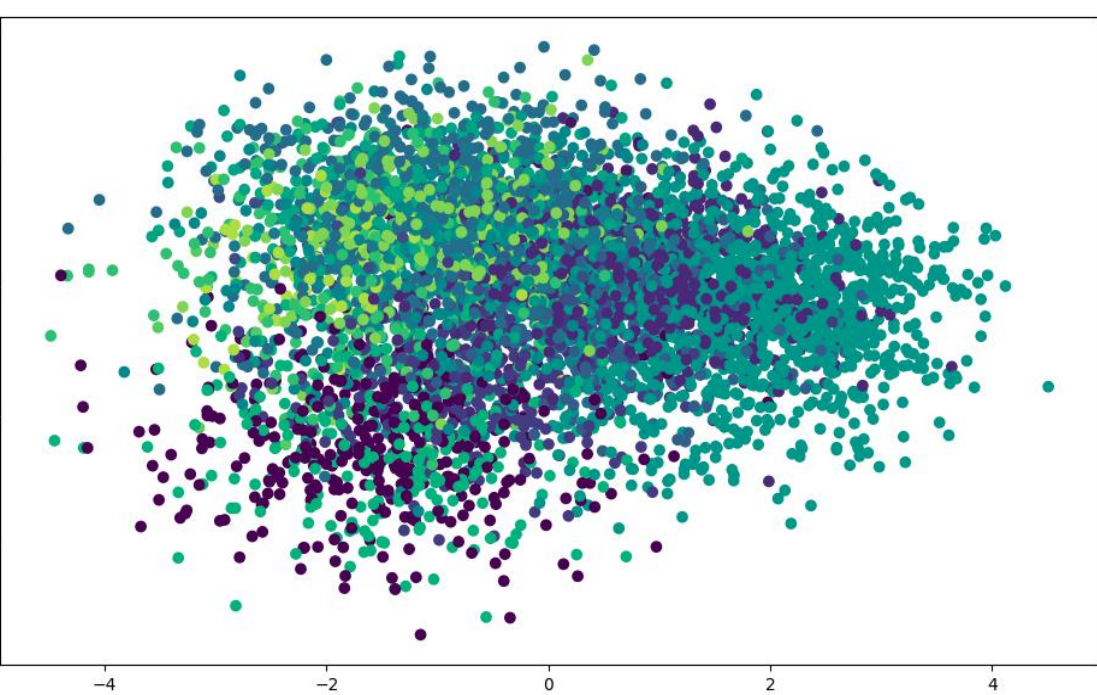
文本向量化处理是文本挖掘、信息检索、情感分析等领域必需的预处理过程，使节点向量包含丰富且有效的语义及结构信息是目前亟待解决的问题。本文首先对科技政策类的文本特征进行分析，分别依照概念与概念间关系的分类体系，用BiLSTM-CRF算法和SVM分别实现对概念与概念关系进行自动标引。并提出了结合推理知识的概念知识网络及进一步融合篇章结构的知识网络构建方法，基于此知识网络模型，本文实现了一种能够融合节点语义、拓扑结构以及类别标签信息的网络表示学习模型，能够充分挖掘并表示文本的语义及结构信息，并通过实验验证了所提方法的有效性。

结合推理知识的TriDNR网络表示学习实验

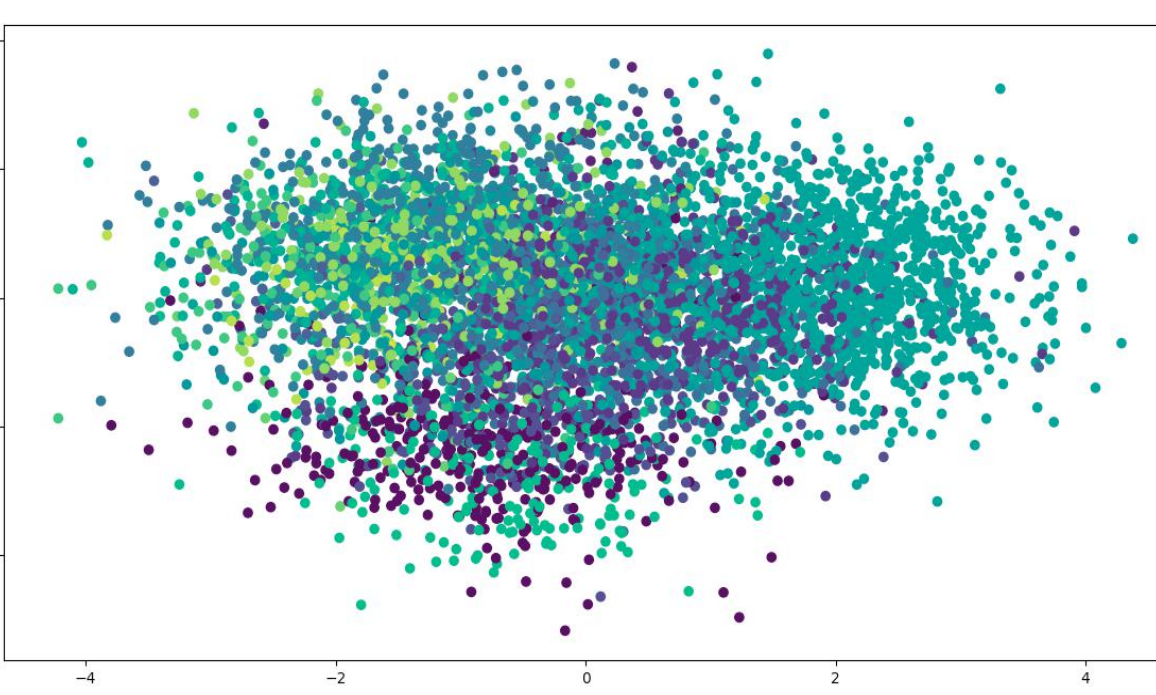
%p	DeepWalk	Doc2vec	DW+Doc2vec	Node2vec	TriDNR	本文方法
10	0.243	0.275	0.341	0.312	0.476	0.512
30	0.315	0.361	0.407	0.386	0.553	0.593
50	0.394	0.428	0.478	0.443	0.618	0.649
70	0.431	0.452	0.501	0.492	0.642	0.683



(a) DeepWalk+Doc2vec向量可视化



(b) TriDNR向量可视化

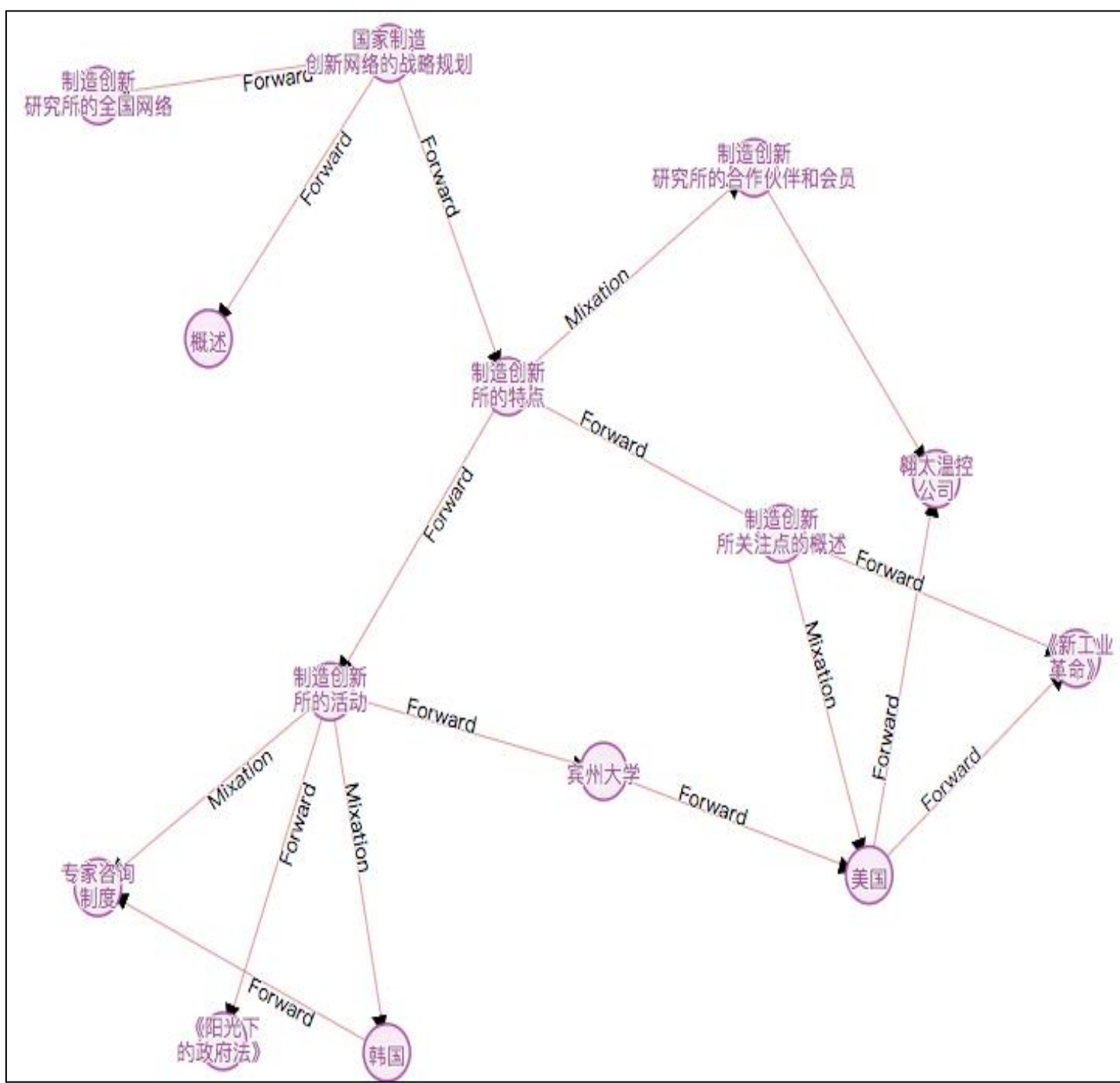


(c) 本文方法向量可视化

TriDNR只考虑了节点之间的拓扑结构信息，没有将节点边的标签信息考虑进去。本文借鉴的知识表示学习模型中的TransE模型，将节点边的五种类别标签Backward、Forward、Mixation、Likelihood和Inclusion融入知识网络中，这五种类别标签揭示了概念之间的推理关系。

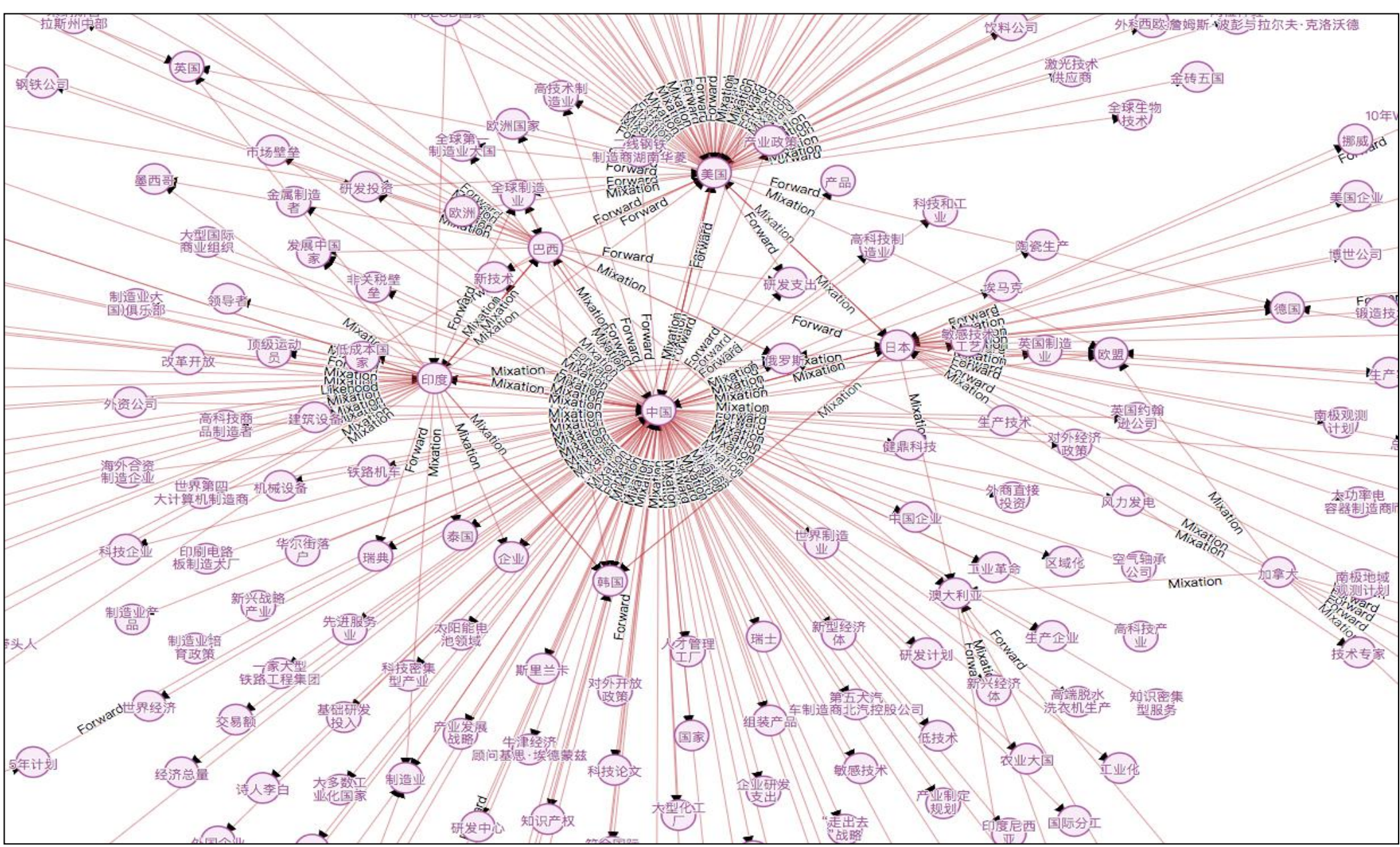
知识表示学习能够同时获取节点表示和边表示，通常被应用在实体链接任务中。由于本文研究只关注将关系标签映射到节点表示中，所以只将节点表示与上一节从节点的拓扑结构、文本语义以及节点标签进行向量相加求平均，最后得到的概念节点向量作为知识网络中每个概念的向量表示。

融合篇章结构的网络表示学习实验



融合篇章结构的知识网络示例图，从图中可以明显地看到树状篇章将知识网络组织起来。不同章节下的相同的概念进行合并，不同概念相互连接，最终形成树状结构与网络结构相结合的融合篇章结构的知识网络模型。

结合推理信息的TriDNR可视化



本文的网络表示学习技术将知识网络的概念节点映射成一个300维的向量表，将其中的拓扑结构信息、语义信息、节点标签信息以及节点之间的推理信息融入到节点表示中，可应用在其他深度学习模型中。

总结

科技政策文本的概念与结构标引为知识网络构建奠定了基础。本文首先定义了知识网络模型，并且分别提出了概念知识网络和带有篇章结构的知识网络模型构建方法。基于知识网络模型，本文通过网络表示学习技术对知识网络中的节点进行表示。对于知识网络中的概念表示，本文首先采用了融合节点语义、拓扑结构以及标签信息的网络表示模型，并在此基础上提出结合知识推理模型的网络表示学习模型改进方法。最终通过SVM分类器和网络表示可视化的方法证明本文改进模型的有效性。对于带篇章结构的知识网络模型的篇章节点表示，本文提出了将篇章节点当作文档，通过Doc2vec算法来获取章节节点表示，最后通过重要节点进行排序的方法验证了篇章节点表示方法的有效性。