

中山大学硕士学位论文

基于消息传递机制的社会关系检测的算法研究 Multiobjective Optimization Algorithms Based on Hybrid Local Search for Multiobjective Pickup and Delivery Problem with Time Windows

学 位 申 请 人: _____ 李雷来

导师姓名及职称: _____ 万海 副教授

专 业 名 称: _____ 软件工程

答辩委员会主席(签名): _____

答辩委员会委员(签名): _____

二零一八年三月十四日

论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：_____

日 期：_____

学位论文使用授权声明

本人完全了解中山大学有关保留、使用学位论文的规定，即：学校有权保留学位论文并向国家主管部门或其指定机构送交论文的电子版和纸质版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆、院系资料室被查阅，有权将学位论文的内容编入有关数据库进行检索，可以采用复印、缩印或其他方法保存学位论文。

学位论文作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

论文题目： 基于消息传递机制的社会关系检测的算法研究

专 业： 软件工程

硕 士 生： 李雷来

指导老师： 万海 副教授

摘 要

摘要内容。

关键词： 关键词

Title: Multiobjective Optimization Algorithms Based on Hybrid Local Search for
Multiobjective Pickup and Delivery Problem with Time Windows

Major: Computer Technology

Name: Xiaoxin Li

Supervisor: Prof. Hai Wan

Abstract

英文摘要

Keywords: 英文关键词

目 录

第 1 章 引言

根据天津大学模板修改的符合中山大学毕业论文（至少是硕士论文）要求的Latex模板。

1.1 使用方法

本模板只包括内容方面的设计预定义，编译自行解决。作者使用的是Windows环境下MikTex+TeXstudio的组合。

1.2 使用建议

1.2.1 普适问题

普遍适用的论文排版问题：

- 图片标题在下，表格在上；一定要有标题，不能只是图1-1；与文字内容的间隔自行把握。
- 参考文献建议使用.bib文件；也有使用Google Scholar的引用的，但有指出当中的“//”不符合规范。
- 部分评审反馈，目录不包含摘要及目录本身，请根据情况自行斟酌。
- 打印时需要右边翻页的问题（每章开始在右边页），可以在生成pdf后通过插入空白页解决（这样插入不会改变页码）；或者尝试设置openright（未测试，有待探讨）。

1.2.2 细节问题

一些细节的问题建议：

- 每个章节都有label，key使用ch:intro形式，以下使用sec:background等。图片key可以参考fig:scenes，表格参考tab:exp。
- 图片、表格尽量在页的顶部，即float优先选择t。



图 1-1 图例

表 1-1 示例表

表头	栏1	栏2	栏3	栏4
内容1	b	—	768 × 576	19
内容1	a	240/7	768 × 576	—

- 另外，为了打印时彩打方便，可以把需要彩打的图片尽量排版在一页，不过比较难调。
- 虽然每个body的tex文件中包含了!Mode:: “TeX:UTF-8”在文件开头，但仍有必要在IDE中将新建的tex文件设为UTF-8编码，否则可能无法正常显示中文。

1.2.3 其他说明

参考文献^[?]目前采用上标表示。使用cite命令。

目前页眉设置：每章第一页页眉只有中间的“中山大学硕士毕业论文”，后续页左边显示“中山大学硕士毕业论文”，右边显示“第n章”。

目前页脚设置：仅包含页码，居中，无横线。

参考文献和附录计算页数，包含在目录，页眉设置同每章第一页。正文前的部分无页眉。

1.3 例子

图例子。label要在caption后。多图或子图方法上网查吧。

表例子。推荐使用这种三行表。缺省值使用三个“-”产生长横线“—”。

公式例子，与普通 Latex 数学公式无异。

$$1 + 1 = 2 \quad (1-1)$$

第 2 章 预备知识

2.1 图像的视觉信息抽取

2.1.1 神经网络

2.1.1.1 一般神经网络

神经网络（neural network）的方面的研究就出现了，早起的神经网络主要是指生物学中的“生物神经网络”，在当前计算机领域特指“神经网络学习”。神经网络最基本的结构是神经元模型，神经元模型如图??，在这个模型中包括输入端、神经元权重（weight）、偏差（bias）、激活函数（activation function）、阈值、输出。在这个模型中，当前神经元接收其它 n 个神经元的输入，与连接权重相乘之后加上偏差，然后激活函数的处理得到激活值输出。

理想中的激活函数是如公式??，将输入值映射为输出值“1”或者“0”，其中“1”对应神经元兴奋，“0”对应于神经元抑制，但是因为该阶跃函数具有不连续、不光滑等性质。因此常采用Sigmoid函数作为激活函数，一般来说激活函数是非线性的、可微的。

$$\text{sgn}(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \quad (2-1)$$

例如sigmoid函数可以把 $\Phi - \infty, +\infty \Psi$ 输入值映射到(0,1)区间内。常见的激活函数还有tanh(hyperbolic tangent, tanh)函数、修正线性单元(rectified linear units, RELU)函数等等。表

表 2-1

函数名称	原函数	一阶导数 $f'(x)$	原函数值域
sigmoid函数	$\sigma(x) = \frac{1}{1+e^{-x}}$	$f^x(x) = \sigma(x)(1 - \sigma(x))$	(0,1)

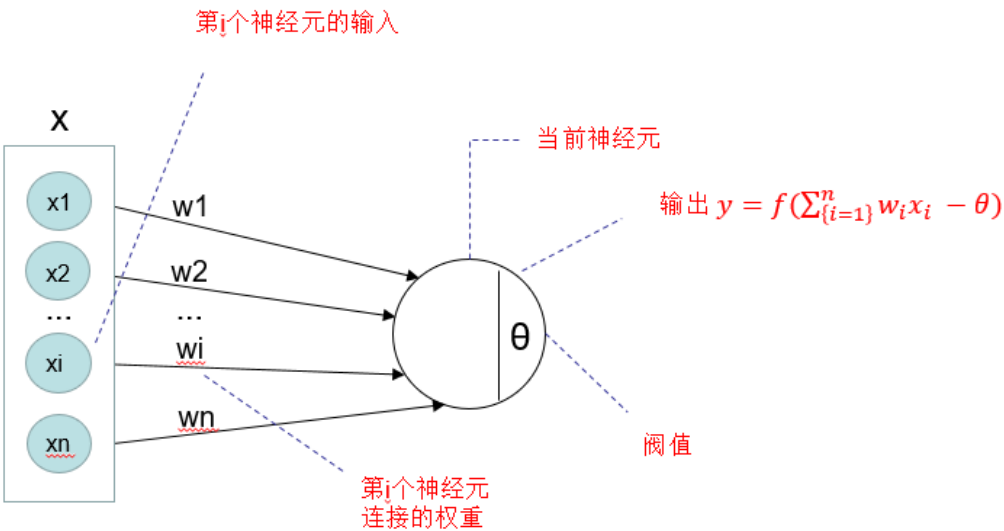


图 2-1 神经元模型示意图

2.1.1.2 卷积神经网络的介绍

卷积神经网络（Convolutional Neural Network,CNN）是深度学习的代表算法之一，由LeCun首次实现并且应用。卷积神经网络的主要作用是提取特征，该网络受到生物学的影响，相比较与全连接神经网络，卷积神经网络的主要特性包括局部感知和参数共享。局部感知指对于具有空间特征的输入来说，每个神经元没必要知道全局的信息，只需要感知局部的信息，然后在更高层将局部的信息合并起来得到更高层的信息。对于权值共享来说，每个卷积核与位置无关，因为假设对于图像来说，其中某一部分的统计特性和其它的部分是一样的，所以对于其中的一个卷积核来说，可以应用到图像上的任何地方去。所以，局部感知和参数共享不仅能提取到更多的特征，并且能大幅度减少参数的数量。因此，卷积神经网络广泛的应用在图像、视频、音频和文本等各种模态的数据上，并且都取得了巨大的成功。

卷积神经网络的特征提取层主要包括两个模块，分别是卷积层（convolutional layer）和池化层（pooling layer），两者的顺序，一般是先通过卷积层，然后是池化层。对于卷积层，主要的作用是提取特征，卷积层的核心是卷积核（kernel），其本质还是神经元。但是卷积核的感受野和全连接的神经元是不同的，这里的感受野是局部的，并且感受野的大小由卷积核的大小控制。如图??所示，当前卷积核的大小是4×4的，对于输入的图片6×6×3，其中图片

输入的3为图片的通道数、 6×6 为高宽，假设滑动的步骤为1，卷积核通过在输入图片上按照步长进行滑动并且进行对应位置的点乘运算，最后形成一个 4×4 的特征图。以上综合起来就是卷积操作，其中 3×3 就是网络的参数。按照惯例，输入的图片可以有固定的高宽和通道数时，卷积核可以有不同的高宽，但是必须是固定的通道数，这里一般和输入的通道数一致。有多少个卷积核，最后就能得到多少个特征图（feature map）。

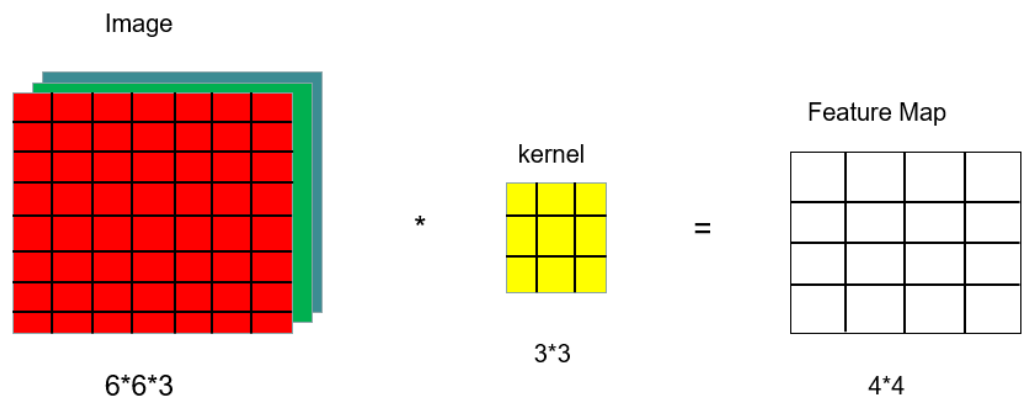


图 2-2 卷积神经网络的卷积层示意图

对于池化层来说，主要的作用是针对卷积层输出的特征图提取主要特征，降低网络的参数，且有防止过拟合的作用。常见的池化包括平均池化(Average pooling)和最大池化(Max pooling)。具体细节如图??所示，池化也是通过类似卷积的操作实现的，在图例中，池化也是以 2×2 在特征图上进行滑动，滑动的步骤为2，而最大池化是选着窗口中的最大值作为输出，平均池化是选择窗口中所有值的平均值进行输出，假设输入的特征图为 $C \times W \times H$ ，那么经过如图例所示的操作后得到的特征图为 $C \times \frac{W}{2} \times \frac{H}{2}$ 。



图 2-3 卷积神经网络的池化层示意图

综合上述对于卷积神经网络的卷积和池化的介绍，因为每层的输入和输

出都表现为特征图的形式，因此卷积神经网络可以和全连接的网络一样可以有多层，并且取得更好的效果。LeNet-5^[2]是Yang LeCun等人在1988年提出的，它是第一个成功应用于数字识别问题的卷积神经网络，在著名的MINIST数据集上，LeNet-5可以取得大约99.2%的准确率。LeNet-5是一个经典的卷积神经网络，前5层分别是卷积层和池化层，后2层全连接层。之后于2012年提出的AlexNet^[2]，其网络结构如图??首次使用Relu激活函数替代Sigmoid，并且验证了其在较深网络上的作用，成功解决了Sigmoid在较深网络的梯度弥散问题，虽然Relu很早就提出了。其次，AlexNet首次在训练中使用dropout层抑制一部分激活的神经元，以避免过拟合，并且通过实践证明了效果。与此同时，模型还采用了数据增强等trick来防止过拟合，使用cuda提高训练速度。而之后提出的VGG^[2]，相比较与之前的LeNet和AlexNet，最大的特点是网络更深，具有16-19层，不包含池化和最后的softmax层。ResNet^[2]是何凯明等人(2016)提出

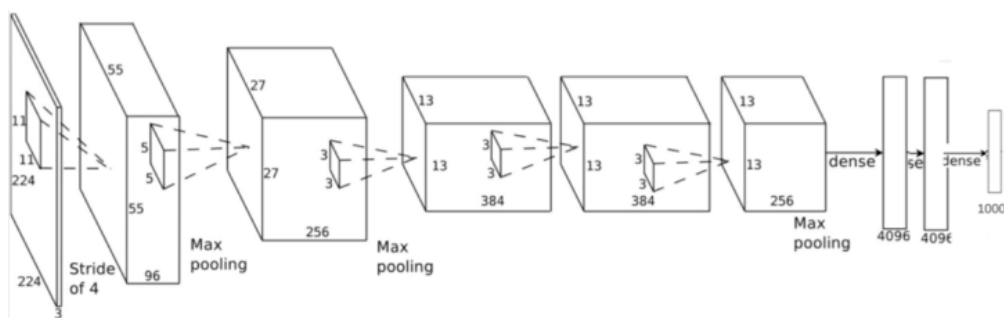


图 2-4 AlexNet网络结构示意图

的，针对前面网络并不能随着层数的叠加而性能的提高，ResNet首次提出了残差学习单元。如图??所示，假设模块的输入为 x ， $F(x)$ 指的是网络中的一系列的张量运算，假设神经网络最优的拟合结果为 $H(x) = F(x) + x$ ，那么神经网络的最优的映射函数 $F(x)$ 为 $H(x)$ 和 x 之间的残差。通过不断的叠加这个模块，可以得到不断加深网络但是不降低网络性能。

综上，卷积神经网络自提出以来，得到了极大的发展，表现为新的激活函数，降低过拟合的dropout层，残差学习模块，加上cuda硬件加速的发展，我们能训练更复杂更深的神经网络，取得更好的性能。

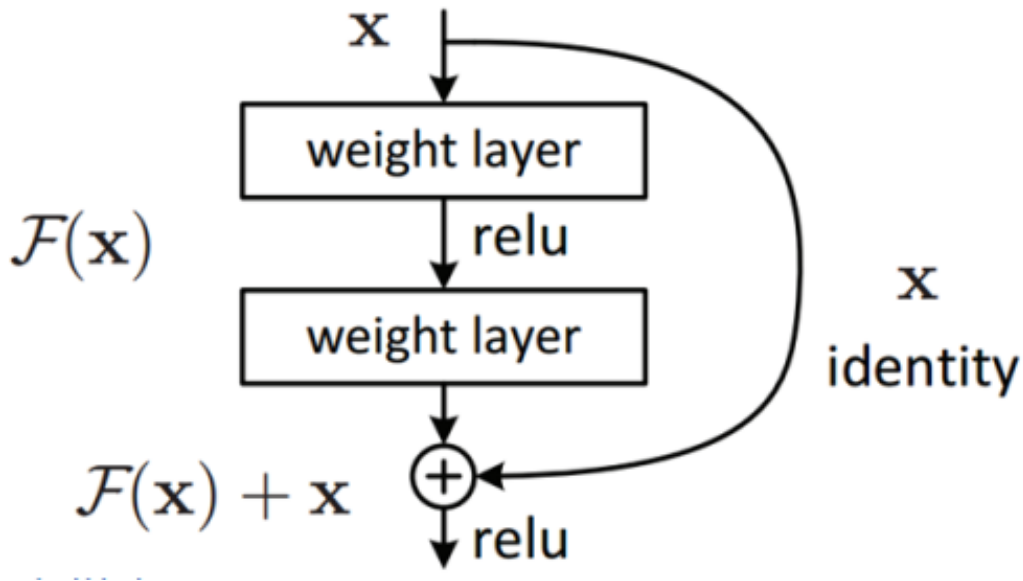


图 2-5 ResNet残差学习单元示意图

2.1.2 物体检测与识别

在社会关系识别的任务中，有一个重要的模块是利用物体识别模型得到人的场景信息，即采用物体识别模型识别去当前图片中包含了哪些物体，得到该物体在图片的区域。物体识别的模型包括Ross等人提出的RCNN^[2]，fast-RCNN^[2]，以及Ren等人（2016）^[2]提出的faster-RCNN。前面提到的三个模型都是基于区域的物体检测模型，RCNN^[2]首次提出在目标图像中有多个目标框，然后判断目标框是否包含物体，具体的检测步骤如下：（1）其中采用选择性搜索的方法得到图片中的所需要的目标框区域，将得到的区域调整为卷积神经网络输入的消息。（2）利用一个预训练好的卷积神经网络，提取第一步得到的区域中的特征。（3）将第二步中得到的特征当作一个线性SVM的输入，得到物体的类别，另外训练一个线性回归模型得到物体的目标框。RCNN的主要缺点是针对一张图片中的每个感兴趣区域，需要遍历提取其中的特征，然后依次执行物体的分类和物体框的回归，需要耗费较多时间。由于全卷积和池化层不改变某个区域在特征图和原图的位置，因此fast-RCNN在RCNN的基础上提出了ROI(region of interest)池化层，将图片输入到卷积神经网络中，对于特征图上的区域，经过ROI池化层进行调整，然后再继续之后的全连接层和一个线性回归层进行分类和目标框的确定。综上，fast-RCNN较大程度上提高了物体检测的性能。由于fast-RCNN在大数

数据集上的表现依然不能满足实际的需求，因为RCNN和fast-RCNN均采用选择性搜索的方法得到所需要的区域，这个步骤是比较耗费时间。因此faster-RCNN提出RPN（region proposal network），RPN主要网络包括两部分，一部分主要是对生成的anchors进行判断是foreground还是background，其中foreground代表目标，另外一部分主要是对检测框的位置进行调整。经过RPN网络后得到候选区域，再利用ROI池化得到特征向量进行物体类别的判断和物体框的进一步精确判断。

综上，以上的篇幅主要是回顾了在社会关系检测的工作中，有用的物体检测方法和一些相关工作。结论是得益于GPU等硬件设备的发展，物体识别领域的算法也得到了快速的发展，尤其是随着特征提取模块的发展，卷积网络越来越深，能学习到更多更丰富的特征。对于一幅图片，我们能在得到更多的、更准确的物体框和类别。

2.2 社会关系检测

本章将回顾社会关系检测领域的一些相关工作，并且对于消息传递机制的介绍，以及消息传递机制的相关工作的一些介绍。

2.2.1 已有工作的简单介绍

社会关系检测是社交网络的一个基础，社会关系检测作为一个重要的多学科问题，在计算机视觉领域受到越来越多的关注。随着这个问题被提出以来，有大量的工作用于从图片中抽取两个人之间的社会关系。主要有Zhang等人（2015）^[2]提出的利用面部表情、年龄、性别、姿势等多种特征的联合模型。Li等人（2017）^[2]提出的多次观察的Dual-glance模型。以及Wang等人（2018）^[2]提出的基于常识知识的深度推理模型GRM。

Zhang等人提出的模型认为从心理学的角度出发，认为人的关系主要由人的面部表情的一些特点决定的。首先，模型设计了一个基准模型用于提取图片中两个人对的特征，对于两个人对，基准模型采用共享参数的深度卷积网络（DCN），利用DCN提取得到的特征分别记为 x^r, x^l ，并且 $\forall x^r, x^l \in \mathbb{R}^{2048 \times 1}$ ，经过一个权重矩阵 $\mathbf{W} \in \mathbb{R}^{4096 \times 256}$ 得到特征向量 x_i 。对于已经标注好的人脸图片两个

人脸分别为 I^l 和 I^r ，利用DCN 提取得到的特征分别记为 x^l 和 x^r ，并且 K^l 和 K^r 。除了图片中本来的特征，模型利用了两张人脸在图片的空间信息。1) 两张人脸的位置分别表示为 $x^l, y^l, w^l, h^l, x^r, y^r, w^r, h^r$ ，其中 x, y 是左上角的坐标， w, h 分别是两个人脸包围盒的宽度和高度。2) 人脸的相对位置 $\frac{x^l - x^r}{w^l}, \frac{y^l - y^r}{h^l}$ 。3) 人脸之间的比例 $\frac{w^l}{w^r}$ 。以上的三项空间特征会和DCN得到的 x_t 拼接来学习得到关系类别。除此之外， w_{gi}, W ,

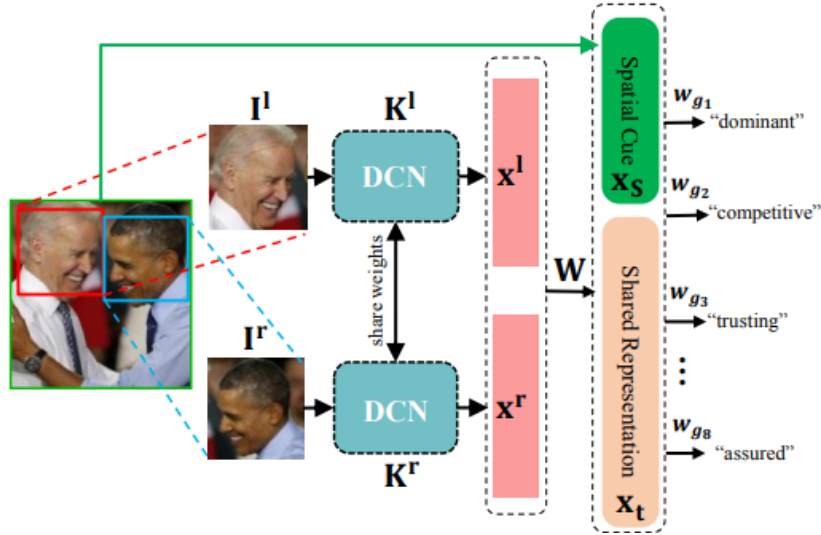


图 2-6 zhang的模型

K^l , and K^r 可以采用标准的正太分布初始化。结合之前符号的定义，该模型损失函数定义如下：

$$\begin{aligned} \arg \max_{\Omega} p(w_{gi=1}^8 | \text{fix } x_t, \text{fix } x_s, \text{fix } I^l, \text{fix } I^r) \propto \\ (\sum_{i=1}^8 p(g_i | x_t, x_s) p(w(g_i))) (\sum_{j=1}^8 K p(k_j^l) p(k_j^r)) p(W), \end{aligned} \quad (2-2)$$

$$s.t. K^r = K^l$$

基于以上的工作，该模型同样认为人的面部属性对最终的关系预测可以起到关键的作用。

Li等人^[7]在前面的工作，针对社会检测的任务提出了包含两个关系粒度的数据集，PISC^[7]。该工作首次提出了利用图片中的场景来协助预测两个人之间的关系，场景具体表示为该图片中的物体。直观来说，如果一幅图片中包含电脑桌子等物体，那么大概率是“同事”关系。模型分为两个模块，first glance和second glance，first glance的输入为一张图片 I 和两个人身体的包围盒。针对图片 I ，首先

修剪出3个小块，前两个小块分别覆盖住两个人， p_1 和 p_2 ，第三个小块覆盖两个人，表示为 p_u 。这三个小块的像素被修正为 224×224 大小，作为后续三个CNNs网络的输入

参考文献

- [1] Wu Y, Lim J, Yang M-H. Online object tracking: A benchmark [C]. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, 2013: 2411–2418.
- [2] Girshick R B, Donahue J, Darrell T, *et al.* Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C/OL]. In Proceedings of CVPR, 2014: 580–587. <https://doi.org/10.1109/CVPR.2014.81>.
- [3] Girshick R B. Fast R-CNN [C/OL]. In Proceedings of ICCV, 2015: 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>.
- [4] Ren S, He K, Girshick R B, *et al.* Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [C/OL]. In Proceedings of NIPS, 2015: 91–99. <http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks>.
- [5] Zhang Z, Luo P, Loy C C, *et al.* Learning Social Relation Traits from Face Images [C/OL]. In Proceedings of ICCV, 2015: 3631–3639. <https://doi.org/10.1109/ICCV.2015.414>.
- [6] Li J, Wong Y, Zhao Q, *et al.* Dual-Glance Model for Deciphering Social Relationships [C/OL]. In Proceedings of ICCV, 2017: 2669–2678. <https://doi.org/10.1109/ICCV.2017.289>.
- [7] Wang Z, Chen T, Ren J S J, *et al.* Deep Reasoning with Knowledge Graph for Social Relationship Understanding [C/OL]. In Proceedings of IJCAI, 2018: 1021–1028. <https://doi.org/10.24963/ijcai.2018/142>.

致 谢

谨此向我的导师张三教授致以衷心的感谢和崇高的敬意！本论文的工作是在张老师的悉心指导下完成的。在传授予我专业知识和宝贵经验的同时，张老师以其严谨的治学态度和精益求精的工作作风不断促进论文相关工作的进行，使我受益匪浅。

在攻读硕士的这三年里，导师和实验室的同学们不仅为我创造了优越的科研和学习环境，使我得以在计算机科学领域中自由翱翔，同时在思想上、人生态度和意志品质方面给予了谆谆教诲，这些教益必将激励着我在今后的人生道路上奋勇向前。特别感谢实验室的甲师兄、乙同学以及其他师弟师妹，他们不仅在学术上给了我许多指引和建议，而且在生活上予以帮助，从他们身上我学到了很多知识。

感谢王五老师及其实验室的同学在领域一、领域二方面的学习给予我的帮助。他们开创性的研究拓展了我的学术视野，无数次的争论和探讨使我的研究工作有了长足的进展。

由衷感谢我的室友A、B和C同学，以及其他经常到我们宿舍进行学习交流的D、E、F和G同学，是他们令我的学习生活都更加充满动力。衷心的感谢我的父母和其他亲朋好友对我的关心、支持和理解，没有他们对我的关心、鼓励和支持，我无法完成现在的硕士学业。

最后，感谢所有曾经教育和帮助过我的所有老师。衷心地感谢为评阅本论文而付出宝贵时间和辛勤劳动的专家和教授们！

李四

二零一四年三月四日