

北京大学学报(自然科学版)
Acta Scientiarum Naturalium Universitatis Pekinensis
doi: 10.13209/j.0479-8023.2018.054

基于句式元学习的 Twitter 分类

闫雷鸣[†] 王超智 严璐琦 贺嘉会 吴宏煜

南京信息工程大学计算机与软件学院, 江苏省网络监控工程中心, 南京 210044;

[†] E-mail: lmyan@nuist.edu.cn

摘要 针对多类别的社交媒体短文本分类准确率较低的问题, 提出一种学习多种句式的元学习方法, 用于改善 Twitter 文本分类性能。该方法将 Twitter 文本聚类为多种句式, 各句式结合原类标签成为多样化的新类别, 从而原分类问题转化为较多类别的 few-shot 学习问题, 并训练深层网络学习句式原型编码。使用多个三分类 Twitter 数据检验本文方法 Meta-CNN, 结果显示该方法的学习策略简单有效, 即便在样本数量不多的情况下, 与传统机器学习分类器和部分深度学习分类方法相比, Meta-CNN 仍然能获得较好的分类准确率和 F1 值。

关键词 元学习; 少次学习; 情感分析; 卷积神经网络

Sentence Style Meta Learning for Twitter Classification

YAN Leiming[†], WANG Chaozhi, YAN Luqi, HE Jiahui, WU Hongyu

School of Computer and Software & Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing 210044; [†] E-mail: lmyan@nuist.edu.cn

Abstract Due to the limited length and freely constructed sentence structures, it is a difficult classification task for short text classification, especially in multi-class classification. An efficient meta learning framework is proposed for twitter classification. The tweets are clustered into many sentence styles corresponding to new class labels. Thus, the original text classification task becomes few-shot learning task. When applying few-shot learning on benchmark datasets, the proposed method Meta-CNN achieves improvement in accuracy and F1 scores on multi-class twitter classification, and outweigh some traditional machine learning methods and a few deep learning approaches.

Key words meta learning; few-shot learning; sentiment analysis; CNN

对于微博、推特这类社交平台的短文本评论信息来说, 在多分类问题上, 即便采用深度学习方法, 分类准确率不高仍然是困扰业界的一个难题。社交平台的文本评论信息字数少、语法格式自由、大量使用缩略语和新词语等, 隐喻、反讽和极性迁移等句型经常出现, 各类型样本数量分布很不平衡, 造成社交平台短文本分类的困难。以 SemEval2017 的 twitter 分类比赛结果为例, 前三名系统虽然在二分类任务(正向、负向)上准确率都超过 86%, 但是对于三分类问题(正向、负向和

中性), 最好系统的准确率仅为 65.8%, F1-score 为 68.5%^[1]。有标签训练样本不足是性能偏低的主要原因之一, 随着分类类别的增加, 样本分布不平衡的情况进一步加剧, 总体需要的训练样本进一步增加。虽然迁移学习策略希望通过迁移到其他领域、利用已有的领域知识来解决目标领域中仅有少量有标记样本的问题^[2], 但是由于社交媒体短文本长度短、形式自由、常违背语法的特点, 难以迁移其他源领域的知识。分类模型泛化能力不足是另一个主要原因, 由于句型的灵活多变, 词

国家自然科学基金(61772281, 61703212, 61602254)资助

收稿日期: 2018-04-21; 修回日期: 2018-08-19; 网络出版时间: 2018-08-22 13:56:10

语的组合形式难以穷尽,训练样本不可能覆盖所有的语义形式,即测试样本中有大量形式没有出现在训练样本中,因此模型无法正确识别。

目前在社交媒体的短文本情感分析方面,特别是多级情感分类方面的研究,仍然面临有标签样本数量不足、分类模型泛化能力不足的挑战。本文提出一种适合少样本、多类别的 twitter 分类框架,该框架基于 few-short learning 策略,利用 deep CNNs 提取样本的 meta-features,用于识别训练样本中未出现过的类型,从而提高分类模型的泛化(generalization)能力。

1 相关研究

词向量被设计成词的低维实数向量,采用无监督学习方法,从海量的文本语料库中训练获得,语法作用相似的词向量之间的距离相对比较近^[3],这就让基于词向量设计的一些模型能够自带平滑功能,为应用于深层网络带来了便利^[4]。一些将词向量与长短期记忆网络(LSTM)相结合的研究都获得明显的性能改善^[5-6]。Kim 等设计的文本卷积神经网络,虽然只有一层卷积层,当其分类性能显著优于普通的机器学习分类算法,例如最大熵、朴素贝叶斯分类和支持向量机等^[7]。Tang 等^[8]基于深度学习设计了 twitter 情感分析系统 Cooolll,将词向量与反映 twitter 文法特点的特征(例如是否大写、情感图标、否定词、标点符号簇集等)进行拼接,以求输入更多有效的特征,在 SemEval 2104 国际语义评测竞赛中,获得第 2 名。深度学习需要大量的训练样本,增加训练样本是非常直接有效的提高分类准确率的方法,但是成本很高、甚至在很多情况下难以实施,这也制约了基于深度学习的文本分类方法的性能。

Few-shot 学习^[9-10]是近年兴起的一种新型元学习技术,它使用较少样本训练深层网络模型,主要应用于图像识别领域,目前只有非常少的文献将其用于文本分析。这种方法首先以 zero-shot(零次)学习和 one-shot(一次)学习出现,逐步发展成 Few-Shot 学习。此类方法的基本思想是,将图片特征和图片注释的语义特征非线性映

射到一个嵌入空间,学习其距离度量。当输入未知样本或未出现在训练集中的新类别样本时,计算样本与其他已知类别的距离,判断其可能的类别标签。虽然有标签的训练样本较少,但此类方法但仍然在图像识别领域(特别是在图片类别达到数百到 1000 的分类任务中)获得成功。Zhang 等^[11]研发了一种基于最大间隔的方法用于学习语义相似嵌入,结合语义相似,用已知类别的样本度量未知类别样本间的相似性。Guo 等^[12]设计了一种新颖的 zero-shot 方法,引入可迁移的具有多样性的样本,并打上伪标签,结合这些迁移样本训练 SVM,实现对未知类别样本的识别。Oriol 等^[13]基于 metric learning 技术和深层网络的注意力机制,提出一种 matching 网络机制,通过支持集学习训练 CNN 网络。Rezende 等^[14]将贝叶斯推理与深层网络的特征表示组合起来,进行 one-shot 学习。Koch 等^[15]训练了两个一模一样的孪生网络进行图像识别,获得良好的效果。一些学者基于“原型”(prototype)概念设计了 Few-Shot 学习方法,但是原型的定义并不统一。Snell 等^[16]提出原型网络概念,将满足 k-近邻的数据对象非线性映射到一个嵌入空间中,该空间中的原型是同类标签样本映射的平均值向量,通过计算未知样本与原型的距离来判别类标签。Blaes 等^[9]定义的全局原型是一种元分类器,希望利用全局特征对图像进行分类。Hecht 等^[17]的研究显示,基于原型的深度学习方法在训练事件和内存开销方面都比普通深度学习方法具有优势。

2 文本句式元学习

受 meta-learning 和图像 Few-Shot 学习的启发,本文提出一种文本句式元学习方法。基本思想为,将多种典型的语句变化视为新的类别和“句式”,即将原本只有几种类别标签的文本样本集合,改造为多种新的类别——“句式”style。划分出更多的类别后,强迫深层模型学习细粒度的语法和语义特征。本文方法包含 4 个基础部分:句式提取、训练片段 episode 构造、句式深层编码和分类模型 Fine-tuning。方法框架如图 1 所示。

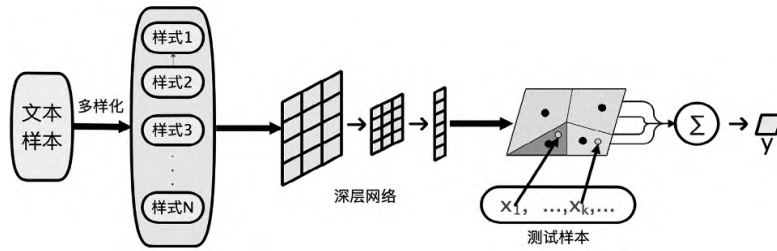


图 1 句式元学习框架

Fig. 1 Sentence styles meta-learning framework

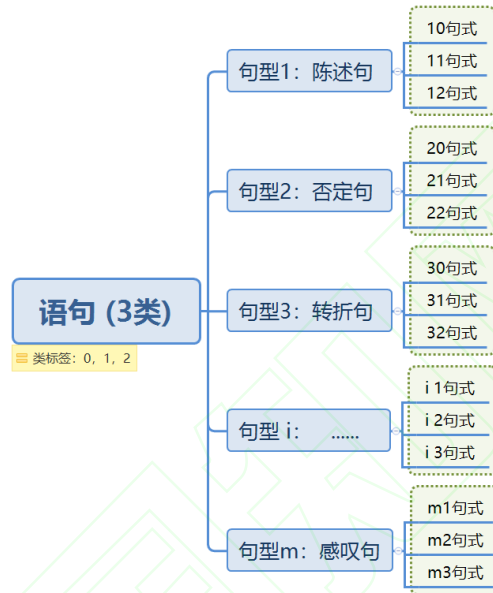


图 2 句式标签划分示意

Fig. 2 Example for sentence style labeling

2.1 提取句式

首先, 将较少类别的文本分类问题转化为较多类别的 Few-Shot 学习问题。本文根据距离相似度, 用 K-均值聚类方法将训练样本划分为若干簇集, 将每个簇集视为一种文本类型, 并进一步划分为句式。

定义 1 句式: 设类标签有 K 种, $L = \{1, 2, \dots, K\}$, 聚类获得的文本类型(句型)有 M 种, $M \gg K$, 不同类型和不同类标签组合为一种新的类别, 称为“句式”(style)。样本集合由原来的 K 种样本, 重新划分为 $N = M \times K$ 种句式, 表示为 $\{s_{ik} | 1 \leq i \leq M, 1 \leq k \leq K\}$, 新的类别标签为 $L' = \{(i, k) | 0 < i \leq N, k \in L\}$, k 为样本原始类标签, 如图 2 所示。训练集中对应新标签的样本称为该句式的支持样本。

定义 2 元句式: 每种句式的支持样本集合的中心样本(即代表样本)称为元句式。元句式可以用样本向量的平均值表示。

新的句式数量大于原来的类别, 相应地, 支持每种句式的有标签样本减少了, 甚至可能有的句式只有一个样本。这类特殊的句式, 可以根据聚类发现的离群点进行添加或删除。我们需要为每种句式构造训练集, 相同句式标签的样本称为该句式的“候选支持集”。将这些样本输入深层网络, 再进行有监督模型训练。需要注意的是, 划分为多种句式后, 导致每种不同句式的支持样本数大大减少。将原分类任务直接转变为支持样本较少的多分类问题, 并不利于提高分类性能。鉴于此, 本文方法借鉴了图像多分类问题的 few-shot 学习思想, 多种句式的划分目的, 不是直接进行多分类学习, 而是用于发现多个具有代表意义的句型原型“prototype”, 通过比较未知类别样本与句型 prototype 的距离, 提高分类准确率。

鉴于缺少有标签的句型样本, 本文采用一种

简单直接的策略, 根据语句相似距离, 用 k -mean 聚类方法提取句式。采用距离相似发现句式的基础是基于词向量模型将语句转化为向量。词向量的优点是可在一定程度上表达语义或语法作用相似, 向量叠加时仍然可以保持原有相似性。因此, 聚类方法不能明确发现否定句、感叹句、隐喻和反讽等实际句型, 但是可以从向量相似的角度将语义和结构上较为相似的样本聚为一类。我们采用 Doc2Vec 模型将语句转化为向量, 将不同长度的语句都转化为相同长度的向量。实现过程如下。

- 1)分词, 训练一个 Doc2Vec 模型, 将每个样本转化为一个向量, 长度为 300。
- 2)设定 k , 调用 k -mean 算法, 对文本向量进行聚类。
- 3)为每个样本分配新的类别编号= 聚类编号 *10+原类别编号; 每种新类别为一种“句式”。
- 4)输出聚类结果。

2.2 训练片段(episode)的构造

在 Few-Shot 学习中, 模型训练过程由多个 episode 构成。 K -shot 学习包含 K 个片段。通常, 对于 N 类“句式”, 每种句式的样本都平均划分为 K 份, 每个 episode 该包含 1 份样本作为训练集, 以及一份测试集样本。为了测试模型对新类别的识别能力, 选择训练集中未出现过的“句式”作为测试集样本。

2.3 元句式深层编码

元句式深层编码即学习句式原型。基本思想是, 将 N 种文本句式的样本向量, 经深层网络(例如 CNN)映射到一个嵌入空间 R^D , 在 R^D 内通过分类算法, 不断调整网络权值, 使得该深层网络根据类别标签和距离, 学习可区分的不同句式的非线性编码。句式原型经深层编码, 被映射到一个非线性空间, 如图 3 所示, 每个区域对应于一种句式原型, 圆点表示该句式的支持样本, 黑色小圆点为该句式的代表点, 即元句式。图中空心圆点表示一个未知标签的新样本经编码进入到嵌入空间中, 可以通过计算到各个原型代表点的距离来判断类标签。

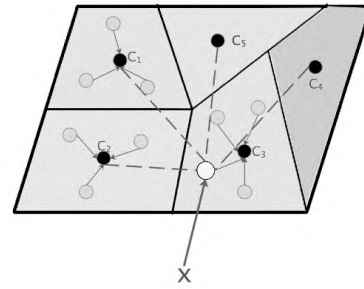


图 3 句式原型与映射空间

Fig. 3 Style prototypes and embedding space

用于编码的深层模型, 采用 CNN 网络构造。基本策略为, 首先用聚类后的、多样化句式的数据有监督地训练网络学习多种句式, 然后使用原始数据优化模型的分类性能, 在已有 CNN 权重的基础上, 训练一个新的 softmax 分类层对原始数据进行分类。

基于 softmax 函数, 分类目标函数可以定义为, 对于未知样本 x^* , 其属于任意类的概率:

$$p(k|x^*, \{P_k\}) = \frac{\exp(-\|h(x^*) - P_k\|^2)}{\sum_{k'} \exp(-\|h(x^*) - P_{k'}\|^2)}$$

根据极大似然假设, 基于交叉熵的损失函数为

$$-\frac{1}{T} \sum_i \log(p(k|x^*, \{P_k\}))$$

综上所述, 本文所提方法属于一种 Few-shot 学习策略, 此类方法可以视为一种元特征学习方法, 侧重特征向量的学习, 发现样本的原型 prototype, 其优化函数通常不以距离为直接目标, 这与 metric-learning 有一定的区别。在实现上也与 metric-learning 有所不同, few-shot 学习需要基于深层网络搭建模型。但是从最新的研究成果^[9-10]来看, 由于 few-shot 学习通常利用 k -近邻思想进行最后的分类, 因此 metric-learning 方法对于 few-shot 有很强的借鉴意义, 二者的融合应该是一种必然的趋势。本文所提“元句式”的概念, 更类似于一种句子“prototype”, 基本思想是发现并深层编码这些基本 prototype, 计算样本与 prototype 样本的距离, 通过加权来判断样本类别。

3 实验

本文使用 3 个公开的 Twitter 数据, 验证本文的方法, 并对结果进行分析。实验服务器配置为 12 核至强 CPU, 256 GB 内存, 8 颗 NVIDIA Tesla K20C GPU, 操作系统为 Ubuntu 14.0。代码基于 Tensorflow 和 Keras, 使用 Python2.7 实现。

本文模型基本结构包括 2 层 1 维卷积层、过滤器 128 个, 过滤器尺寸为 5, 后接 Max-Pooling 层和 Dropout 层, 再接一层全连接的神经网络, 激活函数选择 Relu, 最后是一个 softmax 分类层。参数优化使用 Adam, 交叉熵作为损失函数, batch size 取 50。文本聚类时, 利用 gensim 中的 Doc2Vec 工具实现语句向量化。训练分类模型时, 首先使用聚类后的、增加了句式标签的数据进行模型的预训练, 再使用原始的数据集, 用一个新的 softmax 分类层进行 fine-tuning。

3.1 数据集

1)MultiGames。该数据集为游戏主题的 twitter 数据, 共 12780 条, 由人工进行情感类型标注, 包括正向 3952 条, 负向 915 条和中性 7913 条游戏玩家评论。该数据集由加南大 UNB 大学 Zhang 等的团队发布^[18]。该数据集中的评论多俚语和网络用语以及部分反话。

2)Semeval_b。该数据源自国际语义评测大会 SemEval-2013 发布的比赛数据^[19], 后经不断更新, 所有数据由人工标注为正向、负向和中性 3 种情感类别。由于部分 tweets 的链接失效, 我们共下载 7967 条数据。

3)SS-Tweet。Sentiment Strength Twitter (SS-Tweet)数据集共包含 4242 条人工标注的 tweet 评论。该数据最早由 Thelwall 等^[20]发布, 用于评估基于 SentiStrenth 的情感分析方法。Saif 等^[21]

对该数据重新注释为正向、负向和中性 3 种情感类别。本文实验所用数据包括 1252 条正向、1037 条负向和 1953 条中性。

所有数据集均随机划分为 3 个部分, 验证集和测试集各占 15%, 其余作为训练集。

3.2 实验结果与分析

本文以代价敏感的线性支持向量机为基准方法, 特征提取选择过滤停止词、词性标注(POS)、情感符号 Emoticon 和 Unigram。本文方法命名为 Meta-CNN。用于对比的深度学习方法包括基于自动编码器的 DSC^[18]、文本 Kim-CNN^[7]和一个两层一维卷积层构造的 CNN 模型 2CNN1D。DSC 方法仍然提取了 POS 和 Emoticon 特征, 并过滤停止词, 然后输入自动编码器进行重编码。Kim-CNN 虽然仅包含一层卷积操作, 但在文本分类中常能获得较好的准确率。2CNN1D 的网络结构与本文用于预训练的 CNN 结构相同, 与本文 Meta-CNN 方法进行比较, 用于验证 Meta-CNN 是否能够在双层 CNN 网络基础上改善分类性能。基于 CNN 的方法均不做停止词过滤等预处理, 分词后, 直接使用 Google 的预训练 word2vec 包 GoogleNews-vectors-negative300-SLIM, 转换为词向量构成的语句矩阵, 词向量长度 300, 所有语句样本利用 Padding 操作将长度统一转化为 150 个词, 不足 150 个词时补 0。各方法获得的最佳 Accuracy 如表 1 所示。

表 1 Accuracy 对比(%)

Table 1 Accuracy comparison (%)

数据集	线性 SVM	DSC	Kim-CNN	2CNN1D	Meta-CNN
MultiGames	77.2	78.4	77.6	77.8	91.6
Semeval_b	58.8	63.2	72.1	68.5	89.5
SS-Tweet	66.4	66.4	65.9	62.9	86.6

表 2 正负向样本平均 F1 对比

Table 1 F_1^{PN} Comparison

数据集	线性 SVM	DSC	Kim-CNN	2CNN1D	Meta-CNN
MultiGames	0.435	0.561	0.6	0.571	0.864
Semeval_b	0.554	0.531	0.61	0.567	0.864
SS-Tweet	0.511	0.528	0.679	0.618	0.861

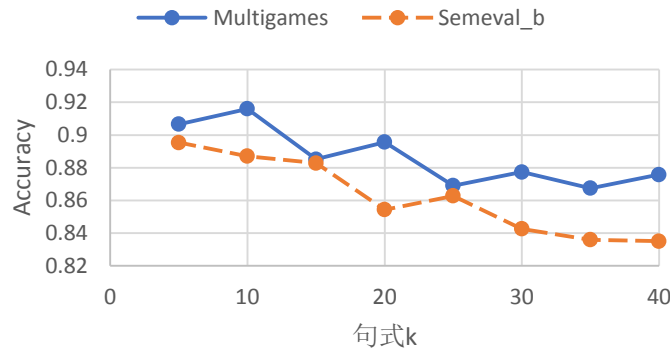


图 4 句式数量 k 对分类性能的影响

Fig. 4 Relationship between style number k and Accuracy

由于数据分布不均衡, 不同类别样本数量有较大差距, 特别是负向标签样本, 通常都比中性标签样本少很多。分布的不平衡性对分类器的 Accuracy 有较大的负面影响。为了更加客观地进行评价, 参照 SemEval 对多分类问题上的评价标准, 我们使用正(Positive)、负(Negative)向样本的平均 F1 值(F_1^{PN})作为多分类任务的评价方法。指标计算方法如下。

$$F_1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}},$$

$$F_1^{PN} = \frac{F_1^P + F_1^N}{2}.$$

各方法 F_1^{PN} 的如表 2 所示。可以看出, 基于深度学习方法的准确率优于线性 SVM。本文提出的 Meta-CNN 方法, 在三个数据集上均取得最高的准确率。与 2CNN1D 分类模型相比, 本文的 Meta-CNN 的准确率远远高于 2CNN1D, 说明本文所提方法中, 在预训练模型的基础上进行调优, 对于改善分类性能是有效的。

样本数量对模型的性能影响明显。SS-Tweet 数据的样本较少, 从 DSC、Kim-CNN 和 2CNN1D 的分类准确率来看, 并未优于线性 SVM。但是本文方法仍然获得较好的分类性能。

句式种类 k 取值对本文方法的准确率有较明显的影响。如图 4 所示。对于数据集 MultiGames, 当句式的聚类数 $k=10$ 时, 获得 91.6% 的准确率。Semeval_b 和 SS-Tweet 数据在 $k=5$ 时获得较优的准确率。随着 k 的增大, 准确率有所波动, 总趋势下降。这是因为, 随着 k 的增大, 分类类别急剧增大, 预训练的模型分类准确率下降, 影响 fine-tuning 时的模型性能。

从实验结果来看, 在句式划分基础上实现的句式原型学习, 在一定程度上改善了分类性能,

说明合理的句式划分有助于提取句子结构特征, 这些特征的引入改善了文本分类性能。但是, 一定句式数量的增加, 导致了类别数量的增加, 显然对分类性能有负面影响。本文基于聚类的句式划分方法并不能对句式进行精确的划分, 因此句式数量越多, 句型特征提取的误差积累越大。合理的句式数量需要通过实验确定。增加训练样本数量是实践中的一种有效提高分类性能的策略。但是, 对于文本分类任务来说, 多少样本数量才是足够的? 目前在理论上并没有明确的结论。从实践和国际上的一些 twitter 分类竞赛结果来看, 数万条训练样本并不能保证获得满意的分类性能, 对于可视为多类别分类的 twitter 情感程度划分任务, 准确率往往只能达到 65% 左右。如果在可承受的成本内, 不能通过数百万条训练样本来训练分类样本, 那么设计少样本学习策略来提升分类器性能, 就成为值得研究的方向。本文就是对少样本的文本分类研究的一种尝试。

4 结语

本文基于元学习和 Few-shot 学习策略, 提出一种文本元学习框架, 通过学习不同的句式特征, 提取更为细粒度的文本语句特征, 以期改善文本分类性能。多个数据集的实验结果证实了本文所提方法的有效性, 对于有标记样本较少情况下的多类别文本分类问题, 使用元学习策略, 可以改善多类别文本分类性能。同时, 本文在“句式”的定义仍旧比较粗糙, 实验结果显示过多的句式数量, 不利于提高分类性能。后续研究方向包括, 改造其他 meta-learning 方法, 使之适用于文本分类任务, 在与本文方法多角度比较中, 改进本文所提方法; 提出更加精细的句式划分策略, 以便准确提取更多的有益语句特征。

参考文献

- [1] Preslav N, Sara R, Svetlana K, et al. Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts. *Language Resources & Evaluation*, 2016, 50(1): 35–65
- [2] 庄福振, 罗平, 何清, 等. 迁移学习研究进展. *软件学报*, 2015, 26(1): 26–39
- [3] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality // *Proceeding of the 27th Annual Conference on Neural Information Processing Systems (NIPS)*. Nevada, 2013: 3111–3119
- [4] 奚雪峰, 周国栋. 面向自然语言处理的深度学习研究. *自动化学报*, 2016, 42(10): 1445–1465
- [5] Cheng J, Zhang X, Li P, et al. Exploring sentiment parsing of microblogging texts for opinion polling on chinese public figures, *Applied Intelligence*, 2016, 45(2): 429–442
- [6] Sundermeyer M, Schlueter R, Ney H. LSTM neural networks for language modeling // *Proceedings of the 13th Annual Conference of the International Speech Communication Association (ISCA)*. Portland, 2012: 194–197
- [7] Kim Y. Convolutional neural networks for sentence classification // *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, 2014: 1746–1751
- [8] Tang D, Wei F, Qin B. Coooolll: a deep learning system for Twitter sentiment classification // *Proceedings of the 8th International Workshop on Semantic Evaluation*. Dublin, 2014: 208–212
- [9] Blaes S, Burwick T. Few-shot learning in deep networks through global prototyping. *Neural Networks the Official Journal of the International Neural Network Society*, 2017, 94: 159–172
- [10] Sachin R, Hugo L. Optimization as a model for few-shot learning // *Proceedings of the 5th International Conference on Learning Representations (ICLR)*. Toulon, 2017: 1–11
- [11] Zhang Z, Saligrama V. Zero-shot learning via semantic similarity embedding // *Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*. Chile, 2015: 4166–4174
- [12] Guo Y, Ding G, Han J, et al. Zero-shot learning with transferred samples. *IEEE Transactions on Image Processing*, 2017, 26(7): 3277–3290
- [13] Oriol V, Charles B, Tim L, et al. Matching networks for one shot learning // *Proceedings of the 30th Annual Conference on Neural Information Processing Systems (NIPS)*. Barcelona, 2016: 3630–3638
- [14] Rezende, D J, Mohamed, S, Danihelka, I, et al. One-shot generalization in deep generative models // *Proceedings of the 33rd International Conference on International Conference on Machine Learning (ICML)*. New York, 2016: 1521–1529
- [15] Gregory K, Richard Z, Ruslan S. Siamese neural networks for one-shot image recognition // *Proceedings of the 32nd International Conference on Machine Learning (ICML)*. Lille, 2015: 1–8
- [16] Jake S, Kevin S, Richard S Z. Prototypical networks for few-shot learning // *Proceedings of the 30th Annual Conference on Neural Information Processing Systems (NIPS)*. Long Beach, 2017: 4080–4090
- [17] Hecht T, Geppert T. Computational advantages of deep prototype-based learning // *Proceedings of 2016 International Conference on Artificial Neural Networks*. Barcelona, 2016: 121–127
- [18] Yan L, Zheng W, Zhang H, et al. Learning discriminative sentiment chunk vectors for twitter sentiment. *Journal of Internet Technology*, 2017, 18(7): 1605–1613
- [19] Nakov P, Rosenthal S, Kiritchenko S, et al. Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts. *Language Resources and Evaluation*, 2016, 50(1): 35–65
- [20] Thelwall M, Buckley K, Paltoglou G. Sentiment strength detection for the social web. *Journal of the Association for Information Science & Technology*, 2012, 63(1): 163–173
- [21] Saif H, Fernández M, He Y, et al. Evaluation datasets for twitter sentiment analysis: a survey and a new dataset, the STS-gold // *Proceedings of the first international workshop on emotion and*

sentiment in social and expressive media:
approaches and perspectives from AI, A workshop
of the XIII International Conference of the Italian

Association for Artificial Intelligence. Turin, 2013:
9-21

