

基于深层学习的词义自动归纳¹

马尔胡甫·曼苏尔, 裴文哲, 常宝宝

北京大学计算语言学教育部重点实验室, 北京, 100871

E-mail: mairgup@gmail.com, williampei1988@126.com, chbb@pku.edu.cn

摘要: 词义自动归纳是自然语言处理领域中的基本任务之一。本文提出了基于词特征向量的上下文特征表示方法, 来探索深层学习方法在词义自动归纳任务中的应用价值。实验结果表明, 使用我们提出的基于词特征向量的上下文特征表示方法, 在没有任何人工特征工程的情况下, 能够达到和现有最好的词义自动归纳效果持平的水平, 说明了作为深层学习方法的副产品的词特征向量在词义自动归纳任务中有使用价值。

关键词: 词义自动归纳; 深度学习;

Deep Learning based Word Sense Induction

Mairgup Mansur, Baobao Chang

Key Laboratory of Computational Linguistics, Ministry of Education

Institute of Computational Linguistics, Peking University, Beijing 100871

E-mail: mairgup@gmail.com, chbb@pku.edu.cn

Abstract: Word Sense Induction is a one of the basic Natural Language Processing tasks. In this paper, we propose a word context representation method based on the distributed word representation, to study the value of Deep Learning in the context of Word Sense Induction. The experimental result shows that, using our Lookup Table based word context representation method, which is totally lack of feature engineering, can get a same level performance with that of the state-of-the-art. This means that, as a byproduct of Deep Learning Methods, distributed word representation can be very useful in Word Sense Induction task.

Keywords: Word Sense Induction; Deep Learning;

1 引言

近几年来, 随着更快的并行计算机系统架构的发展, 人工神经网络又迎来了第二次研究热潮。从原先的浅层的、简单的人工神经网络架构出发, 研究人员进一步发展出了多层神经网络、递归神经网络和深层神经网络等更加复杂的深度学习架构模型。在包括语音识别、图像识别在内的许多机器学习任务中, 深度学习模型打败了许多当前最先进的机器学习方法。

在自然语言处理领域中, 深层学习的直接应用便是基于深层神经网络的语言模型, 这个深层神经网络语言模型凭借对词序列更好的泛化能力, 在困惑度等指标上超越了传统的n-元语言模型以及基于类的n-元语言模型。深层神经网络语言模型训练后的一个副产品就是分布式词特征向量词典, 作为对词的一种特征表示, 我们认为分布式词特征向量可以被应用在很多自然语言处理任务中, 本文就试图将其用在词义自动归纳任务中来探究其应用价值。

用词义来代替词的形式已经被多次证明能够有效提升很多自然语言处理任务的效果。文献[1,2,3]分别在信息抽取、信息检索以及机器翻译这三个领域指出了这种词义代替词的方法的有效性。对于这些等自然语言处理应用而言, 词义往往比词本身更具价值。在早期的研究中, 如想使用词义信息, 往往需要一个人工编辑的词典资源, 如 WordNet 等。

在这些词典资源中, 词典中的词义一般是定义在更广泛意义上的词义, 导致定义本身失去具体的使用环境信息, 并且这类人工编辑的词义资源往往不能反映出目标词在特定的上下文中的具体含义。自然语言一直在不停的进化, 这类词典资源的更新速度很难跟上这个进化速度, 其更新又需要大量的人工标注工作, 人工编辑的代价又非常高昂, 这也使得这些词典资源的应用范围受到了限制。

为了弥补人工编辑的词义资源所存在的上述缺陷, 研究提出了多种无指导的词义自动归纳方法。无指导的词义自动归纳, 就是通过直接从文本数据中自动学习目标词的词义来克服

¹ 本文工作得到自然科学基金项目(61273318)资助, 特此致谢。

人工标注的词义词典的缺陷，而且整个学习过程不需要任何人工标注的资源。

无指导的词义自动归纳任务可以被看成是一个聚类任务。为了正确识别语料中多义词的不同词义，可以依据词的意思将词所在的上下文进行聚类。每一个聚出来的类可以被看作是目标词的一个特定词义。如果将“使用向量来表示词的上下文信息和特征”也视作该任务的一部分的话，词义自动归纳任务可以被形式化的分成两个子任务：

1. 特征表示：选择能够体现出词义差异的特征来表示目标词的上下文信息。
2. 词义聚类：用某种策略将相似的上下文聚成一类，作为一个词义对应的上下文集合。

本文的研究目标是将深层神经语言模型学习到的词的分布式特征表示应用到词义自动归纳任务中来，即将这个词特征表示引入到词义自动归纳任务中的特征表示这一子任务中来。通过这种方式来探究深度学习技术在自然语言处理任务中的应用。

2 分布式词特征表示

作为深度学习架构，深层神经语言模型所学习的目标是一个分布式词特征表示。学习过程结束后，每一个被学习到的词就被表示为一个连续的实数值特征向量，即其分布式特征表示。用特征向量来表示一个词，可以看作是将这个词映射到特征空间中的一个点的过程。功能相似的词，他们所对应应在特征空间中的点之间的距离也是相近的。

这就是说，这个语言模型看到的将不再是一个个独立的词，而是一组组用法上相似的词组成的泛化的类。这也是为什么基于深度学习架构的语言模型能够很好的处理未登录词问题的主要原因所在，即深层神经语言模型具有了“类比”能力，见到训练语料中没有的词序列时，可通过“类比”，给出与和其特征表示相似的词序列的语言模型概率值接近的概率值。

这也是我们之所以看重基于深度学习框架的深层神经语言模型学习到的词的分布式特征表示的原因。因为，当我们学到了这样一种词的特征表示后，在其他的自然语言处理任务中，就可以尽量减少繁复的特征工程，并能有效地降低未登录词对系统性能的影响。

我们将这种词的分布式特征表示形式化的表述为：

对于词典 D 中的每一个词 w ，我们都有一个特征向量 V_w 来描述这个词的所有特征，词典中所有的词的特征向量合起来，组成一个特征向量矩阵 W ，我们称这个特征向量矩阵 W 为特征向量词典（Lookup Table）。

我们将在第三章中介绍如何将这个已学到的特征向量词典应用到词义自动归纳任务中的。通过使用这个特征向量词典，我们可以对目标词所在的上下文进行特征表示，进而对目标词所在的所有上下文按照其特征表示进行自动聚类，聚类结果即为这个目标词的词义类。

3 基于特征向量词典的上下文特征表示方法

本文中，我们设计了两种基于特征向量词典的目标词上下文特征表示方法。根据表示范围的不同，分别被称为基于窗口的上下文特征表示方法和基于卷积的全句特征表示方法。

3.1 基于窗口的上下文特征表示

对于目标词 w 所在的某个上下文 C_w 中的每一个词 w_c ，我们都可以在特征向量词典 W 中找到一个对应的 m 维特征向量 V_{w_c} 来对这个词 w_c 进行特征表示，我们把一个句子中目标词所在的大小为 d 的窗口上下文 C_w 中的所有词的特征向量首尾相接得到一个 $d \cdot m$ 维的上下文特征向量 V_C ：

$$\vec{V}_C = \text{CONCAT}_{w_c \in C_w} \{W \cdot w_c\}$$

用这个向量来表示当前的上下文窗口 C_w 的特征。

3.2 基于卷积的句子特征表示

如前所述，基于窗口的上下文特征表示只考虑到了和目标词临近的局部特征，为了能够对含有目标词的整个句子的特征进行表示，我们设计了一种基于卷积的句子特征表示方法。

对于含有 n 个词的句子而言，我们可以抽取到 $n - d + 1$ 个长度为 d 的小窗口，我们用上述的窗口特征表示方法对这 $n - d + 1$ 个小窗口的特征进行表示，得到一个由 $n - d + 1$ 个特征向量组成的全句特征矩阵：

$$\vec{U} = (\vec{V}_{c_1}, \vec{V}_{c_2}, \dots, \vec{V}_{c_{n-d+1}})$$

我们对这个全句特征矩阵中的特征向量 V_{c_i} 的每一维上取最大值，得到一个压缩后的 $|V_{c_i}|$ 维的特征向量 V_s ：

$$\vec{V}_s[j] = \max_{i=1, \dots, n-d+1} \vec{V}_{c_i}[j]$$

来表示这个句子的全句特征。

3.3 上下文特征表示之间的差异性的度量方法

有了上述的目标词所在上下文特征的向量表示后，我们需要定义上下文之间的相似程度。本文中，我们分别用特征向量之间的距离度量和相似度量来量化上下文之间相似程度。

1. 距离度量，用于衡量两个上下文特征在特征向量空间中的距离，距离越远说明上下文之间的差异越大。

2. 相似度量，即上下文特征之间的相似程度，与距离度量相反，相似度度量的值越小，说明上下文特征之间的相似程度越小，差异越大。

本文分别使用特征向量之间的欧氏距离、特征向量之间的余弦相似度来度量两个特征向量所代表的上下文之间的差异程度。

3.3.1 欧氏距离

欧氏距离 (Euclidean Distance) 是最常见的向量之间的距离度量，衡量的是多维特征空间中各个点之间的绝对距离。公式如下：

$$dist(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

3.3.2 余弦相似度

余弦相似度 (Cosine Similarity) 用特征向量空间中两个特征向量之间的夹角的余弦值作为衡量两个上下文之间差异的大小。相比距离度量，余弦相似度更加注重两个特征向量在方向上的差异，而非距离或长度上的差异。公式如下：

$$sim(X, Y) = \cos \theta = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \cdot \|\vec{y}\|}$$

4 基于 K-means 的词义自动归纳

4.1 聚类算法概述

有了前文所述的上下文特征表示，我们就可以对目标词所在的所有上下文进行聚类了，

聚出来的类就是目标词的词义类。本文中使用了 K-means 聚类算法来对上下文进行聚类。

K-means 聚类算法[4]是聚类任务中经常被使用的一种无指导学习算法。其聚类过程即简单又容易实现，在事先给定类数 K 的情况下可以将给定的数据聚成 K 个类。K-means 聚类算法核心主要由以下几步构成：

- 1.在向量空间选取 K 个点作为初始的类的中心点；
- 2.将每一个聚类单元合并到与其最近的中心点所属的类中；
- 3.当所有的聚类单元被分好类后，重新计算这 K 个类的中心点坐标；
- 4.重复步骤 2 和 3，直到 K 个类的中心点不再变化为止。

在上述算法描述中的第 2 步中，聚类单元被分配到相应的中心点的过程，既可以依据距离度量方式，也可以依据相似度度量方式。当按照距离度量方式分配时，每个聚类单元被分配到与其欧氏距离最近的那个中心点所属的类中；如果是按照相似度度量方式分配，则每个聚类单元被分配到与其余弦相似度最高的那个中心点所属的类中。

用 K-means 聚类算法做词义自动归纳，当 K-means 聚类过程结束时，一个目标词所在的所有上下文被自动聚成了 K 个类，分别代表了目标词的 K 个词义类。而这个词义类个数 K 的指定又是基于 K-means 的词义自动归纳的一个关键性问题，我们将在下一小节中讨论如何合理的指定一个目标词的词义类个数 K。

4.2 意义律与词义类个数 K 的选择

上述的 K-means 聚类方法需要事先给定类的个数 K。对于词义自动归纳任务而言，在事先不知道一个目标词有多少种词义的情况下，我们如何来确定应该聚出的词义类数 K 呢？

解决这一问题的传统方法是分别尝试多种不同的类数 K 来进行聚类，然后选择聚类效果最好的那个 K 值。但是，如果对每一个目标词的聚类过程都用这种办法来选定 K 值，则将消耗大量的计算时间。

为了解决这一难题，我们使用了意义律[5]来指导我们估计这个词义类数 K 的值。意义律告诉我们，一个词的词义种类数和这个词在语料中的出现频率之间存在如下关系：

$$k = af^b \quad (4.1)$$

其中 k 是这个词的词类数，f 是这个词在语料中出现的频率，a 和 b 是相关系数。Edmonds 在[6]中通过实验确定了系数 b 的值为 0.404。

也就是说，只要我们能够知道一组词在语料中出现的频次 f，我们可以仅仅通过选择合适的系数 a 就可以根据公式(4.1)估算出这些词分别可能存在的词义类数 k。与传统的一个个试 K 值的方法相比，这种定义词义类个数 K 的方法效率更高。

从公式(4.1)中可以看出，系数 a 可以影响聚出来的类的粒度。较大的系数 a 会产生更多更细化的词义类，每个词义类中的上下文更具有相关性，聚类效果也会更好。基于意义律的词义类数 K 值的确定方法，使我们可以通过调整系数 a 的值来满足不同的聚类粒度需求。

5 实验

5.1 实验设置

5.1.1 实验语料

作为实验语料，我们使用 Semeval-2010 词义自动归纳和词义消歧评测[7]的标准评测语料，该评测要求针对 100 个目标词进行自动词义归纳，包括 50 个名词和 50 个动词。评测语料包括训练语料和测试语料两部分，训练语料包含 879,807 个含有目标词的句子实例，测试集中包含 8,915 个句子实例。平均词义类个数为 3.79。

我们在本文中用于词义自动归纳任务的特征向量词典，来自Collobert等人在[8]中实现的SENNA系统，其特征向量词典是在英文维基百科上用其深层神经网络语言模型通过无指导的方法训练得到的，训练过程共耗时两个月的时间。其中，词典D包含130,000个词，对词典中的每一个词 w ，都有一个50维的特征向量来表示其特征信息。

为了使用意义律来定义一个词的词义类个数，我们需要知道这些词在一份大语料上出现的频次。我们在英国国家语料库 (British National Corpus) 中统计了每个目标词出现的频次。该英国国家语料库是包括新闻语料和口语语料在内的含有一千万词的综合语料库资源。

5.1.2 评价方法

Manandhar 等人在[7]中提出了两种评价方法，分别是无指导评价方法和有指导的评价方法。无指导的评价方法通过 V-Measure[9]和 Paired F-Score[10]两种评分标准来评价词义自动归纳系统得到的词义类的归纳效果。

有指导的评价方法中，词义自动归纳系统给出的结果被随机的分成两个部分(两种分法：一种按照 8:2 进行划分，另一种按照 6:4 进行划分)，其中较大的一部分被称为映射部分 (Mapping Part)，用来将自动归纳出来的词义类映射到标准答案中的词义类上，映射后用于在第二部分中计算相应的准确率与召回率。具体的映射规则参见[7]。为了防止不同的划分方法带来的评价结果的波动，该评价的结果是在每一种划分情况下，基于 5 中随机的划分映射所得到的评价结果的平均值。

在 Semeval-2010 评测中，两种无指导的评价方法都没有能够很好的区分出所有参加评测的系统的效果好坏。从[7]中的各个系统的不同评价结果中可以看到，V-Measure 评价方法更偏好聚出的类数更多的参评系统，即词义自动归纳系统归纳出来的词义类数越大则相应的 V-Measure 评分越高，这样一来，对于那些归纳能力强但是聚出来的词义类数少的参评系统是相对不公平的；而用 Paired F-Score 评价方法来评价时，基于高频词义标注法 (MFS) 的基准系统结果优于所有的参评系统。

因此，本文中我们选择了更具有实际意义的有指导的评价方法来作为我们的词义自动归纳方法的效果的评价标准。

5.2 实验结果及分析

5.2.1 对意义律参数的调整实验

在我们的基于 K-means 聚类方法的词义自动归纳系统中，需要调整的参数只有一个，那就是意义律中用来确定类数 K 时的参数 a 。我们针对不同的 a 值进行了多组词义自动归纳实验，表 5.1 中给出了在意义律参数 a 取不同的值时词义自动归纳系统的实验结果。

在本实验中，我们使用了基于特征向量词典的窗口上下文特征表示方法，窗口大小为 6，聚类过程中使用了欧氏距离作为上下文的差异度量方法。

从表 5.1 中可以看出，随着意义律系数 a 的变化，词义自动归纳系统的效果会有小的波动，但基本保持在 0.62 以上的准确率和召回率的水平。在随后的实验中，我们对意义律系数 a 的值不再进行调整，都取 0.255 为意义律系数 a 的值。

表 5.1 对意义律参数 a 的调整实验结果

参数 a	准确率(8:2)	召回率(8:2)	准确率(6:4)	准确率(6:4)
0.240	0.6266	0.6202	0.6144	0.6060
0.255	0.6300	0.6244	0.6154	0.6074
0.270	0.6284	0.6218	0.6186	0.6090

0.285	0.6294	0.6222	0.6192	0.6086
0.300	0.6313	0.6230	0.6246	0.6124
0.330	0.6298	0.6210	0.6240	0.6112

5.2.2 相似度度量与距离度量的对比实验

在相同的特征表示下，聚类方法中使用不同的上下文特征差异的度量方法会达到不同的聚类效果。本文中我们分别使用了欧氏距离和余弦相似度来表示上下文特征向量之间的差异，并进行了相似度度量和距离度量之间的对比实验。

本实验分两组进行，其中一组的聚类过程中使用了欧氏距离作为上下文特征差异度量方式，另一组使用了余弦相似度作为上下文特征差异的度量，两组实验都使用了基于特征向量词典的窗口上下文特征表示方法，窗口大小为 6，意义律系数 α 都取 0.255。

在表 5.2 中给出了这两组实验的结果。从表中的结果可以看出，基于欧氏距离的度量方式能达到更好的词义聚类效果。

表 5.2 相似度度量与距离度量的对比实验

	准确率(8:2)	召回率(8:2)	准确率(6:4)	准确率(6:4)
欧氏距离	0.6300	0.6244	0.6154	0.6074
余弦相似度	0.6240	0.6156	0.6140	0.6014

这个结果可以这么理解：特征向量词典本质是用一个向量来表示一个词，我们使用这个特征向量词典来表示窗口上下文特征，其实质是相当于把窗口上下文映射到了特征向量空间中的一个点，这些点的坐标就是其对应窗口上下文的特征向量，在这个映射过程中，类似的窗口上下文所对应的点之间的绝对距离会小些，即在特征向量空间上更近一些，对应到上下文差异的度量上来讲，就是欧氏距离更能体现出上下文之间差异，会有更好的聚类效果。

5.2.3 特征窗口大小选择实验

我们在使用基于特征向量词典的窗口上下文特征表示方法时，为了探究不同的上下文窗口大小对词义自动归纳任务效果的影响，我们分别在不同的窗口大小设置下做了对比实验。

这三组实验的其他实验设置均相同，唯一不同的就是上下文窗口的大小。三组实验均使用了欧氏距离作为上下文差异的度量方式，意义律参数 α 取 0.255。第一组的窗口大小为 4，即考虑目标词前后各两个词；第二组的窗口大小为 6，即考虑目标词前后各三个词；第三组的窗口大小为 10，即考虑目标词前后各五个词；这三组实验分别代表了在小窗口、中等窗口、大窗口的上下文特征表示下的词义自动归纳效果。

表 5.3 中给出了不同的特征窗口大小的选择下词义自动归纳实验的结果。可以看出，过小或者过大的窗口选择都不会达到太好的词义自动归纳效果，说明对目标词词义有决定性因素的词往往出现在目标词周围不太大的窗口范围内，而过小的窗口范围往往捕捉不到决定目标词词义的相关词条，而过大的窗口范围会导致过多的噪音被带入进来，效果都不好。

表 5.3 特征窗口大小的选择实验

	准确率(8:2)	召回率(8:2)	准确率(6:4)	准确率(6:4)
Window=4	0.5838	0.5770	0.5704	0.5608
Window=6	0.6300	0.6244	0.6154	0.6074
Window=10	0.6100	0.6066	0.5998	0.5938

5.2.4 基于卷积的句子特征表示的实验

前文中，我们介绍了另一种利用特征向量词典来对上下文特征进行表示的方法，即基于

卷积的句子特征表示方法。和基于窗口的特征表示方法不同，这种表示方法考虑了全句范围的上下文特征，即把目标词所在的句子看作是目标词的上下文来进行特征表示。

为了探究这种全局的特征表示方法是否会对词义自动归纳系统带来好处，我们进行了两组基于卷积的句子特征表示法的词义自动归纳实验。我们分别取小窗口的大小为 3 和 5 的情况下进行相应的词义自动归纳实验，实验中使用了欧氏距离为上下文特征差异度量方法，意义律系数 α 的值继续取 0.255。

表 5.4 中给出了这两组基于卷积的句子特征表示方法的实验结果。

表 5.4 基于卷积的句子特征表示实验

	准确率(8:2)	召回率(8:2)	准确率(6:4)	准确率(6:4)
Window=3	0.5988	0.5894	0.5866	0.5760
Window=5	0.5962	0.5880	0.5816	0.5702

从表 5.4 中可以看出，小窗口大小的选择对于词义归纳效果的影响不是很大，但是和基于窗口的特征表示方法相比，这种基于卷积的句子特征表示方法并不能带来词义自动归纳效果的提升，这一点又一次证明了我们前面得到的结论，即决定目标词词义的词条往往出现在目标词周围不大的窗口范围内，因此使用合理窗口大小的窗口上下文特征表示方法会得到更好的词义自动归纳效果。

5.2.5 与 Semeval-2010 评测结果的比较

表 5.5 中给出了我们的实验结果同 Semeval-2010 评测中的部分结果之间的对比情况。由于评测结果中没有提供相应的准确率与召回率数据，我们只比较这几个系统的 F 值结果。

表 5.5 与 Semeval-2010 评测结果的比较结果

	F 值(8:2)	F 值(6:4)
UoY	0.6240	0.6196
Duluth	0.6050	0.5950
MFS	0.5870	0.5830
Random	0.5730	0.5650
我们的方法	0.6244	0.6074

其中，参评系统 UoY 使用了基于图聚类的词义自动归纳方法[11]，是本次评测中成绩最好的系统；Duluth 使用了基于 K-means 的词义自动归纳方法[12]，是本次评测中使用 K-means 聚类方法的系统中结果最好的一个；MFS 是指使用目标词的最频繁词义来标注这个目标词的所有句子实例所得到的结果；Random 是指多轮随机划分词义类后的平均结果；我们的方法指的是基于窗口的上下文表示法在使用欧氏距离的情况下所得到的结果。

与我们的词义自动归纳方法相比，UoY 和 Duluth 都使用了大量的特征工程，不仅使用了一阶共现词特征，还用了二阶、三阶等更复杂的特征，他们的方法很大程度上依赖于这样的大规模特征工程，并涉及从这些特征中选择有效特征和过滤无效特征等繁复的预处理过程。

和他们的方法相比，我们的方法几乎没有涉及任何繁复的特征工程。我们使用的词特征向量词典是在维基百科语料上通过无指导的训练过程学习到的，我们的上下文特征表示只是简单的使用上下文中的词对应的特征表示向量来对其进行特征表示，整个过程完全可以看作是没有人工参与的过程。这也是我们的方法优于 UoY 系统和 Duluth 系统所用的方法之处。

虽然，我们的方法所得到的实验结果没能超过 Semeval-2010 评测中的最好结果，但是，通过使用几乎没有特征工程的简单方法达到的词义自动归纳效果，与使用了繁复特征工程后所得到的评测最好结果持平，这在一定程度上说明了我们所提出的基于词特征向量词典的词义自动归纳方法的有效性。也充分说明了词的分布式特征表示对自然语言处理任务的重要性。

6 结论

多年来,自然语言处理领域的核心研究目标就是让机器学习知识的表示。深度学习方法的提出,让基于人工神经网络的知识表示学习方法又一次获得了生机。深度学习,即让计算机深度学习分布式特征表示的方法,在近二十年来的人工智能相关领域引起了研究风暴。很多基于这种方法的深度学习框架被提出,并在许多相关的自然语言处理任务中得到了应用。

本文中,我们在词义自动归纳任务中提出了一种使用分布式特征词典的词义自动归纳方法,通过使用这种几乎没有特征工程的简单方法,达到的词义自动归纳效果,与使用了繁复特征工程后所得到的 Semeval-2010 评测最好结果持平,说明分布式特征表示对于自然语言处理任务是有价值的。

参 考 文 献

- [1] Joyce Yue Chai, Alan W Biermann. The use of word sense disambiguation and information extraction system[J]. AACL/IJCNLP. 1999:850–855
- [2] Ozlem Uzuner, Boris Katz, Deniz Yuret, et al. Word sense disambiguation for information retrieval[C]. PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE. JOHN WILEY & SONS LTD 1999, 985–986
- [3] David Vickrey, Luke Biewald, Marc Teysier, Daphne Koller. Word-sense disambiguation for machine translation[C]. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. ACL, 2005, 771–778
- [4] James MacQueen, et al. Some methods for classification and analysis of multivariate observations[C]. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. California, USA, 1967, vol. 1, 14
- [5] George Kingsley Zipf. The meaning-frequency relationship of words[J]. The Journal of General Psychology. 1945, 33(2):251–256
- [6] Philip Edmonds. Disambiguation, lexical[J]. Encyclopedia of Language and Linguistics Second Edition Elsevier. 2006
- [7] Suresh Manandhar, Ioannis P Klapaftis, Dmitriy Dligach, Sameer S Pradhan. SemEval-2010 task 14: Word sense induction & disambiguation[C]. Proceedings of the 5th International Workshop on Semantic Evaluation. Association for Computational Linguistics, 2010, 63–68
- [8] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, Pavel Kuksa. Natural language processing (almost) from scratch[J]. The Journal of Machine Learning Research. 2011, 12:2493 – 2537
- [9] Andrew Rosenberg, Julia Hirschberg. V-measure: A conditional entropy-based external cluster evaluation measure[C]. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2007, vol. 410, 420
- [10] Javier Ariles, Enrique Amigó, Julio Gonzalo. The role of named entities in web people search[C]. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2. Association for Computational Linguistics, 2009, 534–542
- [11] Ioannis Korkontzelos, Suresh Manandhar. UoY: Graphs of Unambiguous Vertices for Word Sense Induction and Disambiguation[C]. Proceedings of the 5th International Workshop on Semantic Evaluation. Uppsala, Sweden: ACL, 2010, 355–358.
- [12] Ted Pedersen. Duluth-WSI: SenseClusters Applied to the Sense Induction Task of SemEval-2[C]. Proceedings of the 5th International Workshop on Semantic Evaluation. Uppsala, Sweden: Association for Computational Linguistics, 2010, 363–366.