

第 1 次印刷的勘误表

1. P26, 公式 3.6 下面的一行中 $w_1 + w_2 + \dots + w_V = 1$ 应为: $w_1 + w_2 + \dots + w_n = 1$ 。
即等式左边最后一项 w_V 应修改为 w_n 。
2. P30, 公式 3.12 中 w_{1-n+1} 应修改为: w_{i-n+1} 。
3. P33, 最后一行公式 3.33 中第 2 个 Σ 下面第变量 $w_i \in V$ 应修改为 $w_i \in V_{neg}$, 并在该公式下面增加一句说明: “其中, $w_i \in V_{neg}$ 为负样本集合。”
4. P38, 公式 3.49 中等式右边 $\max_{k=1}^l$ 应该为: $\max_{k=1}^d$ 。
5. P39, 公式 3.51 中等式右边 $\max_{k=1}^l$ 应该为: $\max_{k=1}^d$ 。
6. P62, 第 9 行: “…作为优化目标对优化目标求导置零,” 应修改为: “…作为优化目标。对优化目标求导置零, ”。
7. P62, 公式 4.31 上面一行最后一句: “常常对 θ_i 进行拉普拉斯平滑” 应修改为: “常常对 $\theta_{i|j}$ 进行拉普拉斯平滑”。
8. P69, 公式 4.49 中右边括号中 $\hat{y}_{k_i} - y_{k_j}$ 应改为: $\hat{y}_{kj} - y_{kj}$ 。即下标 j 与 k 是平级的, y 不再加黑。
9. P83, 5.2 节第 5-6 行: “以下从样本间的相似性和簇间的相似性两个方面, 分别介绍文本相似性度量方法。” 修改为: “以下从样本间的相似性、簇间的相似性和样本与簇之间的相似性三个方面, 分别介绍文本相似性度量方法。”
10. P84, 第 1-2 行, 删除 “以下从样本间的相似性、簇间的相似性和样本与簇之间的相似性三个方面, 分别介绍文本相似性度量方法。”。
11. P84, 倒数第 8 行和第 291 页 “名词术语索引” 中英文单词 “simility” 应修改为: “similarity”。
12. P86, 公式 5.8 应修改为: $J(x_i, x_j) = \frac{|x_i \cap x_j|}{|x_i \cup x_j|}$ 。即将原式中 x 的上标改为下标, 去掉括号。该页中第 5.2.2 节的第一段, 按同样方式修改 x 的上标。
注意: 公式中的 x 是加黑的。
13. P87-88 中的公式 5.14、5.15 和 5.16, 以及算法 5.1 中, 字母变量 S 均应修改为: C 。即公式 5.14 改为: $\arg \min_C \sum_{k=1}^K \sum_{x \in C_k} \bullet$; 公式 5.15 修改为: $\arg \min_{C^{(t)}} \sum_{k=1}^K \sum_{x \in C_k^{(t)}} \bullet$;

5.16 公式右边修改为： $\frac{1}{|C_k^{(t)}|} \sum_{x_i \in C_k^{(t)}} \bullet$ 。注意：公式中的 \mathbf{x} 都是加黑的。

14. P88, 算法 5.1 中第 2 行“输出”那一行和第 8 行, 字母 S 均改为 C , 不加粗。
15. P91, 算法 5.2 中的“输出”行和第 1、7、8、10 行字母 S 均换成 C 。
16. P97, 倒数第 5 行中 D 集合内 d_2 后面多了一个逗号, 应删除。
17. P98, 第 2 行, d_i 和 d_j 中的变量 d 应加粗。
18. P99, 第一个公式等式右边分子上的 Σ 下面, $d \in C_m, d \neq d'$ 应修改为: $d' \in C_m, d' \neq d$ 。
19. P101, 第 3 行中, “多义性 (polysems) 和同义性 (synonymys)” 英文拼写有误, 应修改为: “多义性 (polysemy) 和同义性 (synonymy)”。
20. P102, 6.1 节倒数第 4 行 Darid 应修改为: David。
21. P104, 第 6.2.3 节第一句“利用 SVD 将项-文档矩阵分解之后”修改为: “利用 SVD 将词项-文档矩阵分解之后”。
22. P106, 6.3 节第 2 和第 3 行, 两处 Hoffmann 均修改为 Hofmann。
23. P109, 表 6.2 第 4 行, “先验分布超参数 (K 维向量)” 中删除 “(K 维向量)”, 第 5 行中, “先验分布超参数 (V 维向量)” 中删除 “(V 维向量)”。
24. P165, 最后一行, Z =原公式的分母, 即删除分子部分。
25. P177, 公式 9.7 中等号两边的 n 应修改为 T ; 公式 9.8 的右边括弧内: i_{t+1} 应修改为: y_{t+1} 。
26. P197, 公式 9.47 的右边, $SR(m_i \rightarrow m_j)$ 应修改为: $SR(m_j \rightarrow m_i)$ 。
27. P207, 第 1 行“ e_2 之前的第二个词语”应修改为: “ e_2 之后的第二个词语”; 该页最后一行: “是否为空上述例子中 CPHBO 取值为空”应修改为: “否则为空, 上述例子中 CPHBO 取值为空”。
28. P238, 10.3.1 句子压缩方法的第一段第 4 行, “字串”应修改为: “子串”。
29. P241, “(5) 模型解码”下面的第一个句子的句号改为逗号。
30. P254, 公式(10.43)下面第一行, W_{ij}^{encn} 应为: W^{encn} 。另外, 该页上提到的矩阵变量 W 均应加黑。

以下参考文献缺少页码范围或者页码等信息有误, 按如下说明修改:

- [1] P266, [Bing et al., 2015], 原页码有误, 更改为: pp. 1587-1597.
- [2] P268, [Ding et al., 2019], 增加页码: pp. 6343-6350.
- [3] P268, [Do et al., 2012], 论文集名称有误, 页码没问题, 论文集名称“IJCNLP and COLING”更改为: “EMNLP-CoNLL”。

- [4] P268, [Dumais et al., 1998], 增加页码范围: pp. 281-285.
- [5] P269, [Firth, 1957], 原页码范围有误, 更改为: pp. 1-32.
- [6] [Fleischman and Hovy, 2004], 原论文集名称有误, 而且缺少页码范围, 更改为: In *Proceedings of the Conference on Reference Resolution and Its Application*, pp. 1-8.
- [7] P269, [Gamon, 2004], 原页码范围缺少上界, 更改为: pp. 841-847.
- [8] P270, [Hu and Liu, 2004], 原页码上界有误, 更改为: pp. 168-177.
- [9] P271, [Kim and Hovy, 2004], 原页码范围缺少上界, 更改为: pp. 1367-1373.
- [10] P273, [Li et al., 2010b], 论文集名称有误, 更改为: In *Proceedings of COLING*, 即去掉原 “/ACL”。
- [11] P273, [Liao and Veeramachaneni, 2009], 论文集名称有误, 页码范围没问题。论文集名称 “NAACL-HLT” 更改为: “In *Proceedings of the NAACL-HLT 2009 Workshop on Semi-supervised Learning for Natural Language Processing*”。
- [12] P273, [Lin, 2004], 原页码范围有误, 更改为: pp. 74-81.
- [13] P275, [McCallum and Li, 2003], 论文集名称有误, 页码范围没问题。论文集名称 “NAACL-HLT” 更改为: “In *Proceedings of 7th CoNLL at HLT-NAACL*”。
- [14] P276, [Miwa and Bansal, 2016], 原页码有误, 更改为: pp. 1105-1116.
- [15] P276, [Narasimhan et al., 2016], 原页码有误, 更改为: pp. 2355-2365.
- [16] P276, [Nema et al., 2017], 原页码有误, 更改为: pp. 1063-1072.
- [17] P277, [Pang and Lee, 2004], 原页码范围缺少上界, 更改为: pp. 271-278.
- [18] P277, [Peng et al., 2017], 原页码有误, 更改为: pp. 101-115.
- [19] P282, [Xia and Ding, 2019], 增加页码: pp. 1003-1012.
- [20] P282, [Xia et al., 2013a], 原页码范围下界有误, 更改为: pp. 2176-2182.
- [21] P282, [Xia et al., 2019], 增加页码: pp. 5285-5291.
- [22] P284, [Zheng et al., 2017], 原页码有误, 更改为: pp. 1227-1236.