基于深度学习的情感分类

Ke Wang

Institute of Computer Science and Technology
Peking University

May 30, 2018

Contact: wangke17@pku.edu.cn

Outline

1. 情感分类概述

1.1 简介

1.2 基本模型

2. 近期进展

2.1 利用数据信息

2.2 可解释性

2.3 新模型

2.4 新任务

3. 未来趋势和方向

Outline

简介

情感分类

目的是确定文本的情感极性（例如正面或负面等）。

- 作用
  - 商品口碑分析
  - 商品评论分析
  - 网民舆情监控
  - 股票预测 [1]
  - ... ...

简介

- 领域
  - 产品评论 [2, 3, 4, 5, 6, 7, 8]
  - 电影评论 [9, 2, 10, 5]
  - 社交媒体 [11, 12, 13]
  - 新闻文章 [14]
  - 论文评审 [15, 16]
  - ... ...
- 任务
  - 文档级的情感分类
  - 句子级别情感分类
  - Aspect-level 情感分类
  - 情绪分类，讽刺/幽默检测，跨模态情感分类，...

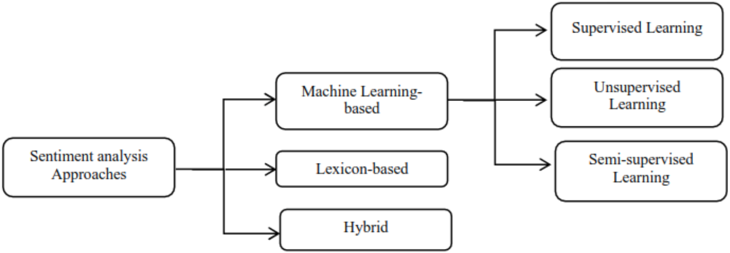
基本模型

- 传统方法
  - Lexicon-based

    基于 POS 和 WordNet [2, 17, 18]
  - 监督方法

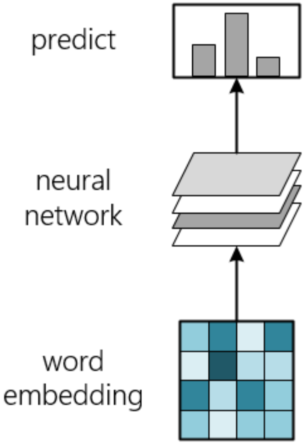
    如支持向量机、最大熵、朴素贝叶斯等，和特征组合。
  - 无监督方法

    使用情感词典、语法分析和句法模式的不同方法。
  - 一些综述: [19, 20]



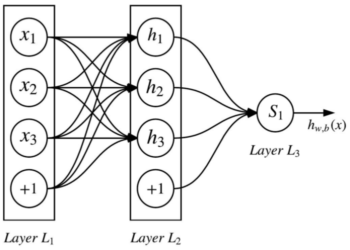
基本模型

- 深度学习方法
  - Word2Vec 词嵌入
    将高维 one-hot 形式表示的单词映射成低维向量，有 skip-gram 和 CBOW 方法等。
  - 分类模型
    DNN， CNN, RNN，LSTM 等
  - 预测
    用 Softmax 函数等预测最终类别



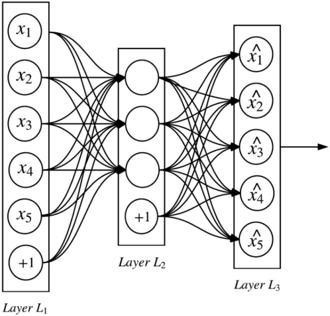
基本模型

- 前馈神经网络
  - 特点
    不考虑输入数据可能具备的任何特定结构。
  - 优点
    简单，有效，可以学习非线性表征。
  - 缺点
    大量隐藏层难以训练，参数很多，容易梯度消失和过拟合。（通常与正则化一起使用）



基本模型

- 自编码器与降噪自编码器 [21, 22]
  - 特点
    一个三层神经网络，目标是使输出值近似等价于输入值。
  - 优点
    提取更鲁棒的特征
    可以无监督训练自编码器
    中间表征可以初始化网络
  - 缺点
    训练困难。



基本模型

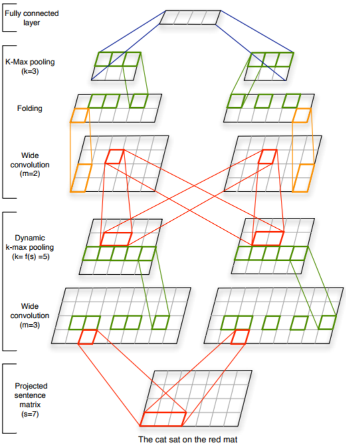
- 卷积神经网络 [5, 23]
  - 特点

    每一个单元都只会和上一层部分单元相连接。一般由若干个卷积和池化操作组成。
  - 优点

    参数大大减少。

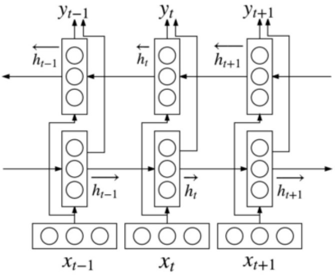
    提取抽象程度更高的特征。
  - 缺点

    没有考虑文本里特有的序列特征。



基本模型

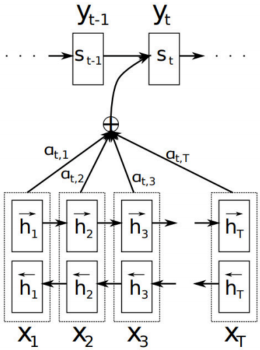
- 循环神经网络 [24]
  - 特点
    一条单向流动的信息流是从输入单元到达隐藏单元，同时另一条从隐藏单元到达输出单元。变种：GRU，LSTM[24]，Bi-RNN[25]。
  - 优点
    参数较少
    擅长于处理时序特征
  - 缺点
    容易误差累积，梯度消失或消失爆炸 [26]



Outline

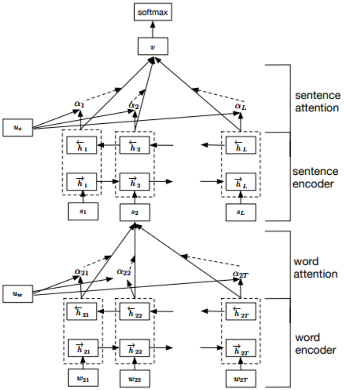
利用数据信息

- 注意力机制
  - 特点
    最早用于机器翻译中， 用于处理时序数据的长期依赖性问题。[27]
  - 优点
    能处理时序数据的长期依赖性问题



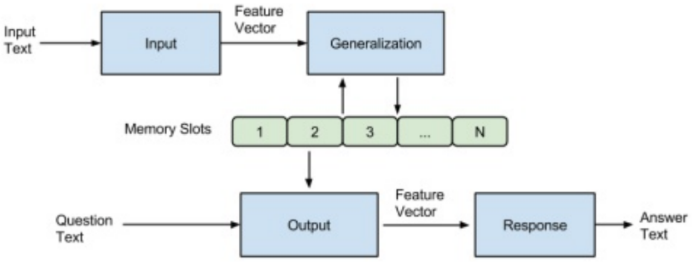
利用数据信息

- 注意力机制
  - 层次式的 Attention [28, 7, 29]
  - 融合各种信息计算
    - Topic 信息 [30]
    - Aspect 信息 [31]
    - 语言特征 (POS, character, ...)[29]
    - ... ...



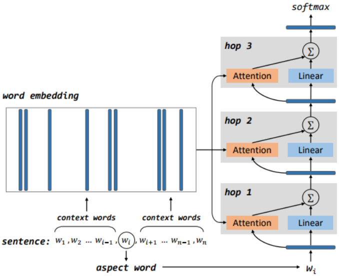
利用数据信息

- 记忆网络
  - 特点
    模型主要包含一系列的记忆单元（可以看成是一个数组，每个元素保存一句话的记忆）和 I，G，O，R 四个模块。[32]
  - 优点
    使用外部存储器模块来保存数据，比传统模型有更多的记忆空间。



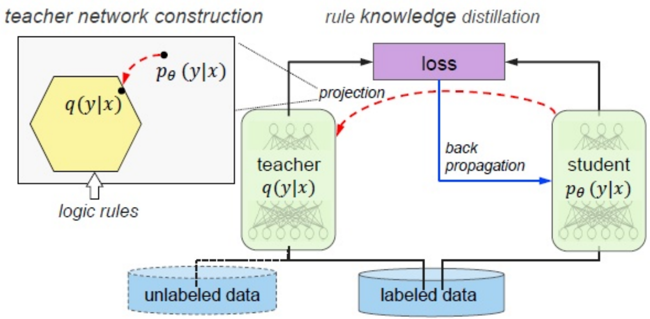
利用数据信息

- 记忆网络
  - Aspect Level Sentiment Classification with Deep Memory Network [33]
    - hop 代表层数, 参数共享
    - 输入是 aspect 词向量
    - memory 中存放的是当前分析的一句话中所有的词语所对应的词向量
  - 未看到更多后续工作。



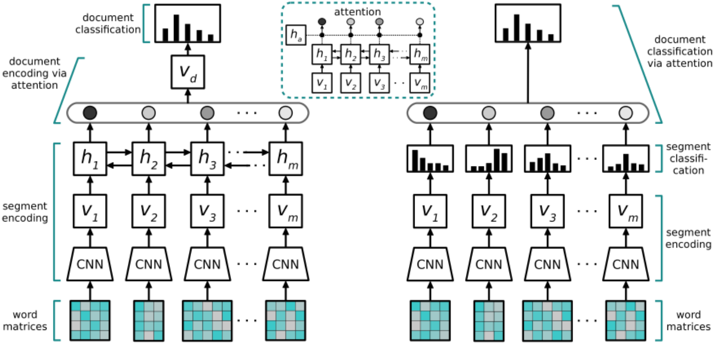
可解释性

- 引入逻辑知识 [34]
  - 特点
    - 教师网络使用软逻辑来编码一阶逻辑信息
    - 教师网络的逻辑信息转移到学生网络
    - 学生网络有更好的泛化能力
  - 优点
    - 帮助训练，提高模型的可解释性。



可解释性

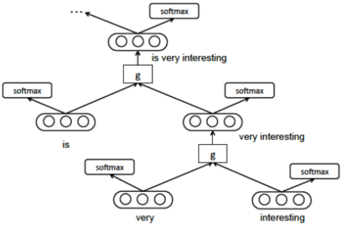
- 多示例学习 [35]
  - 特点
    - 训练分类器的 instance 是没有 label 的，但是 bags 却是有 label 的。
  - 优点
    - 能预测 instance 的 label, 提高了模型预测的可解释性。

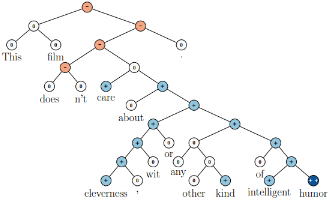


(a) HIERNET (b) MILNET

新模型

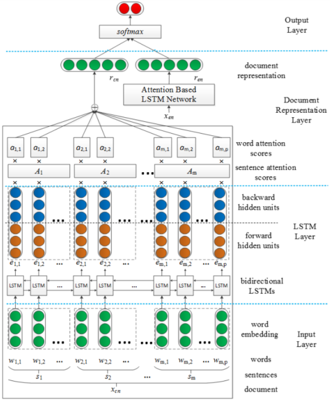
- 递归神经网络 [36, 37]
  - 特点
    句子的语法树中的左右子节点通过一层线性神经网络结合起来，根节点的这层神经网络的参数就表示整句句子。
  - 优点
    利用上句子语法树信息





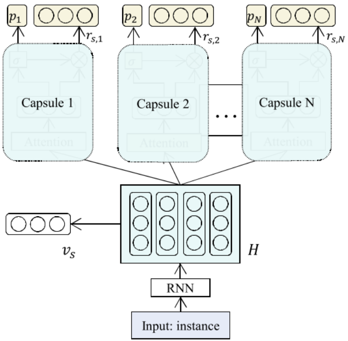
新模型

- 跨语言情感分析 [7, 38]
  - 特点
    用机器翻译工具将语料贫乏的语言翻译成语料丰富的语言（英语），然后再利用新语言的词嵌入，情感词典等等。
  - 优点
    - 利用语料丰富的语言帮助语料匮乏的语言



新模型

- 胶囊网络 [39]
  - 特点
    神经元的输出从一个标量变成了一组向量。有很多个胶囊组成，胶囊之间使用囊间动态路由算法通信。
  - 优点
    - 能够学习到数据之间的相对关系
    - 相比于 CNN，它只需要学习一小部分数据
  - 应用
    - RNN-Capsule [40]
    - 每个胶囊专注于一个情感类别



新任务

- 新的文本类型
  - 论文同行评议 [15, 16]
  - 讽刺分析 [41, 42]
  - 幽默检测 [43]
  - ... ...
- 新的数据类型
  - 文本 + 声音数据 [44]
  - 文本 + 图像数据 [45]
  - ... ...

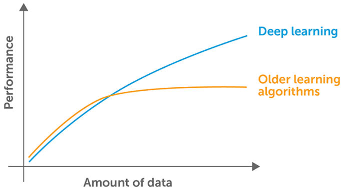
This paper presents low-rank bilinear pooling that uses Hadamard product. The paper implements · · ·
· · · · · ·
I like the insights about low-rank bilinear pooling using Hadamard product presented in the paper. *However, it could not be justified that low-rank bilinear pooling leads to better performance than compact bilinear pooling. It does lead to reduction in number of parameters but it is justification of why low-rank bilinear pooling is better than other forms of pooling.*

Prediction: Accept
Summary:
[+0.19] I like the insights about low-rank bilinear pooling leads· · ·
[+0.12] The paper presents new insights into element-wise · · ·
[+0.06] The paper presents a new model for the task of VQA · · ·
[+· · ·] · · ·
[- 0.12] it could not be experimentally verified that low-rank · · ·
[- 0.11] I would like the authors to provide experimental · · ·
[- 0.05] It is not very clear from reading the paper.
[-· · ·] · · ·

Outline

未来趋势和方向

- A New Model？
  - 用更少的数据量来达到同样的效果
  - 融合更多先验知识
  - 可解释
- A New Technology？
  - 具有更强的泛化性，不依赖于训练数据
  - 具有通用性，不依赖于数据类型
- A New Theory？
  - 什么是情感？依赖于主体而存在？



How do data science techniques scale with amount of data?


Diolch
Kiitos
umesc
Sheun
Dekuju/Dekujeme
Kasih
Mamnoon
Todah
Shnorhakalutiun
Teᵒekkür
Hvala
Shokriya
Dziękuję
Shokrun
Spaas
Mul
Ačí
Gamsahapnida
Ngiyabonga
Xie
Dank
Takk
Cam
Gra
or
Dhanyavaad
Dakujem
Waad
Dhanyavadaalu
Kop
Salamat
Merci
al
Dankie
Go
Grazie
Faleminderit
Daw
Dhanyavad
Khopjai
Kruthagnathalu
krap
ThankYou
kun
Arigatou
Tack
Shukriya
Or
Dhonnobaad
ederim
Asante
Grazzi
raibh
Gracias
Nandree
Blagodariya
Gomapsupnida
Euxaristo
Kun
Hain
Dhan
Fyrir
Terima
Enkosi
Danke
dank
daa

References I

[1] M. P. Baker and J. Wurgler, "Investor sentiment in the stock market," *Journal of Economic Perspectives*, vol. 21, no. 2, pp. 129–152, 2007.

[2] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *ACL*, 2002, pp. 417–424.

[3] K. Dave, D. M. Pennock, and D. M. Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews," in *International Conference on World Wide Web*, 2003, pp. 519–528.

[4] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification." in *ACL*, 2007, pp. 187–205.

[5] Y. Kim, "Convolutional neural networks for sentence classification," *empirical methods in natural language processing*, pp. 1746–1751, 2014.

[6] D. Tang, B. Qin, and T. Liu, "Learning semantic representations of users and products for document level sentiment classification," in *Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, 2015, pp. 1014–1023.

[7] X. Zhou, X. Wan, and J. Xiao, "Attention-based lstm network for cross-lingual sentiment classification," in *EMNLP*, 2016, pp. 247–256.

[8] N. Li, S. Zhai, Z. Zhang, and B. Liu, "Structural correspondence learning for cross-lingual sentiment classification with one-to-many mappings," *national conference on artificial intelligence*, pp. 3490–3496, 2016.

[9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," *empirical methods in natural language processing*, pp. 79–86, 2002.

[10] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," pp. 1631–1642, 2013.

References II

[11] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, “Target-dependent twitter sentiment classification,” in *Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, pp. 151–160.

[12] Y. Ren, Y. Zhang, M. Zhang, and D. Ji, “Context-sensitive twitter sentiment classification using neural network,” pp. 215–221, 01 2016.

[13] J. Deriu, A. Lucchi, V. De Luca, A. Severyn, S. Müller, M. Cieliebak, T. Hofmann, and M. Jaggi, “Leveraging large amounts of weakly supervised data for multi-language sentiment classification,” ser. WWW ’17, Republic and Canton of Geneva, Switzerland, 2017, pp. 1045–1052.

[14] N. Godbole, M. Srinivasaiah, and S. Skiena, “Large-scale sentiment analysis for news and blogs,” in *International Conference on Weblogs and Social Media*, 2007.

[15] D. Kang, W. Ammar, B. Dalvi, M. Van Zuylen, S. Kohlmeier, E. Hovy, and R. Schwartz, “A dataset of peer reviews (peerread): Collection, insights and nlp applications,” 2018.

[16] K. Wang and X. Wan, “Sentiment analysis of peer review texts for scholarly papers,” *SIGIR*, 2018.

[17] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, “Lexicon-based methods for sentiment analysis,” *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, 2011.

[18] M. K. Dalal and M. A. Zaveri, “Opinion mining from online user reviews using fuzzy linguistic hedges,” *soft computing*, vol. 2014, p. 2, 2014.

[19] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 12, pp. 1–135, 2008.

[20] B. Liu, “Sentiment analysis and opinion mining,” 2012.

[21] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy layer-wise training of deep networks,” pp. 153–160, 2007.

[22] G. E. Hinton and R. Salakhutdinov, “Reducing the dimensionality of data with neural networks.” *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

References III

[23] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, “A convolutional neural network for modelling sentences,” *meeting of the association for computational linguistics*, pp. 655–665, 2014.

[24] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[25] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[26] Y. Bengio, P. Y. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[27] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” *international conference on learning representations*, 2015.

[28] Z. Yang, D. Yang, C. Dyer, X. He, A. J. Smola, and E. H. Hovy, “Hierarchical attention networks for document classification,” pp. 1480–1489, 2016.

[29] F. Yang, A. Mukherjee, and E. Dragut, “Satirical news detection and analysis using attention mechanism and linguistic features,” *ACL*, 2017.

[30] H. Chen, M. Sun, C. Tu, Y. Lin, and Z. Liu, “Neural sentiment classification with user and product attention,” in *Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 1650–1659.

[31] W. Che, Y. Zhao, H. Guo, Z. Su, and T. Liu, “Sentence compression for aspect-based sentiment analysis,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2111–2124, 2015.

[32] J. Weston, S. Chopra, and A. Bordes, “Memory networks,” *CoRR*, vol. abs/1410.3916, 2014. [Online]. Available: http://arxiv.org/abs/1410.3916

[33] D. Tang, B. Qin, and T. Liu, “Aspect level sentiment classification with deep memory network,” *empirical methods in natural language processing*, pp. 214–224, 2016.

[34] Z. Hu, X. Ma, Z. Liu, E. Hovy, and E. Xing, “Harnessing deep neural networks with logic rules,” pp. 2410–2420, 2016.

References IV

[35] S. Angelidis and M. Lapata, “Multiple instance learning networks for fine-grained sentiment analysis,” 2017.

[36] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, “Recursive deep models for semantic compositionality over a sentiment treebank,” 2013.

[37] Q. Qian, B. Tian, M. Huang, Y. Liu, X. Zhu, and X. Zhu, “Learning tag embeddings and tag-specific composition functions in recursive neural network,” in *Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, 2015, pp. 1365–1374.

[38] P. Singhal and P. Bhattacharyya, “Borrow a little from your rich cousin: Using embeddings and polarities of english words for multilingual sentiment classification,” in *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, 2016, pp. 3053–3062. [Online]. Available: http://aclweb.org/anthology/C/C16/C16-1287.pdf

[39] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” 2017.

[40] J. H. Y. L. Yequan Wang, Aixin Sun and X. Zhu, “Sentiment analysis by capsules.”

[41] M. Zhang, Y. Zhang, and G. Fu, “Tweet sarcasm detection using deep neural network,” in *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, 2016, pp. 2449–2460. [Online]. Available: http://aclweb.org/anthology/C/C16/C16-1231.pdf

[42] L. Peled and R. Reichart, “Sarcasm SIGN: interpreting sarcasm with sentiment based monolingual machine translation,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, 2017, pp. 1690–1700. [Online]. Available: https://doi.org/10.18653/v1/P17-1155

[43] D. Yang, A. Lavie, C. Dyer, and E. Hovy, “Humor recognition and humor anchor extraction,” in *Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 2367–2376.

References V

[44] D. Bertero, F. B. Siddique, C. S. Wu, Y. Wan, R. H. Y. Chan, and P. Fung, “Real-time speech emotion and sentiment recognition for interactive dialogue systems,” in *Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 1042–1047.

[45] S. Poria, E. Cambria, D. Hazarika, N. Majumder, A. Zadeh, and L. P. Morency, “Context-dependent sentiment analysis in user-generated videos,” in *Meeting of the Association for Computational Linguistics*, 2017, pp. 873–883.