

万物皆可嵌入

沈志立¹⁾

¹⁾南京大学 计算机科学与技术系, 南京市 210023

摘 要 合身的衣物不仅能够降低用户因尺寸问题而产生的退换货成本, 而且能够提高用户的满意度, 因此根据用户身材为其建议合适的尺寸对于电商平台具有重要意义。本文首次提出了 SiZeLet, 该网络从用户, 衣物和评价三个维度提取相关离散属性值的嵌入表征, 并将其与数值属性值相结合以预测衣物对于用户来说是否合身。该模型在 Product Fit 数据集中验证集上的 F1 分数达到了 0.713 ± 0.01 , 并通过消融实验证实了嵌入表征的有效性。同时, 本文以此为契机, 对 fastText 与 GLoVe, textCNN 和 BiRNN 等 NLP 领域技术进行了对比。最后, 本文的相关代码将会开源在 <https://github.com/ZhiliShen/SiZeLet>。

关键词 深度学习; 神经网络; 自然语言理解

Everything Can Be Embedded

SHEN Zhi-Li¹⁾

¹⁾Department of Computer Science & Technology, Nanjing University, Nanjing

Abstract A Fitting clothing can not only reduce the return cost of users due to size problems, but also improve user satisfaction. Therefore, suggesting suitable sizes for users based on their body shapes is important for e-commerce platforms. This paper proposes SiZeLet for the first time. The network extracts embedding representations of relevant discrete attribute values from the three dimensions of user, clothing, and review, and combines it with continuous attribute values to predict whether clothing fits the user. The F1 score of the model on the validation set of the Product Fit dataset has reached 0.713 ± 0.01 , and the effectiveness of the embedding representation was confirmed through ablation experiments. At the same time, this article uses this as an opportunity to compare fastText with GLoVe, textCNN and BiRNN and other NLP technologies. Finally, the relevant code of this article will be open sourced at <https://github.com/ZhiliShen/SiZeLet>.

Key words deep learning; neural networks; natural language processing

1 引言

电商平台的兴起改变了人们的购物习惯, 越来越多的人选择在网上购买商品, 其中就包括衣物。但用户在线上购买衣物无法进行试穿或搭配, 这也导致了因为尺寸不合身而产生的退换货, 这无疑增加了客户的时间成本与电商平台的经济成本。因此, 根据用户与衣物的相关属性为其推荐合身的尺寸是一个具有挑战性与现实意义的课题。

深度学习已经在多个领域展现了其卓越的有效性: 例如图像识别, 目标检测和自然语言理解等^[1]。但是在本课题中, 用户与衣物均具有大量的离

散属性: 例如用户身材类型, 衣物种类, 租借目的等, 其中用户标识符属性 (user id) 与衣物标识符属性 (item id) 均属于极度稀疏的离散属性, 例如: 在 Product Fit 数据集中衣物标识符属性种类有 5850 之多, 而用户标识符属性甚至多达 105571。若对这些离散属性使用独热编码表示, 无疑会带来“维度灾难”^[2], 同时, 独热编码之间无法表示属性内不同种类之间的相对关系, 导致深度神经网络无法捕捉数据中有效的特征以进行合理的分类, 因此这些属性的加入很难提升深度神经网络的性能与泛化能力。

但我们在数据处理中也不能将用户标识符属性或衣物标识符属性等稀疏离散属性弃之不用, 因

为衣物的合身与否与用户自身和衣物本身有着极强的相关性,例如:某一用户因其肩长过宽,导致对于其他用户合身的衣物对其不合身,但是这样的特征是无法通过用户的体重,胸围和身高等属性捕捉,只能通过用户标识符进行关联;又例如:某一衣物由于设计或生产工艺问题,导致其码数偏离于正常衣物,这样的特征也是无法从衣物中的租借目的或衣物种类等离散属性中提取得到,也只能通过衣物标识符对其进行捕捉,因此这些稀疏离散属性又必须得到合理的利用。

本文则从词嵌入(word2vec)^[3]中得到启发,word2vec摒弃了独热向量的表示方法,将每个词映射为一个定长的向量,同时这些向量也能较好地表达词与词之间的相似关系,这极大地提升了下游自然语言处理任务的性能。而这与本课题的任务不谋而合,每个词都可以视为词汇这个属性下的一个实例,同样的,每个衣物标识符可以视为衣物标识符这个属性下的一个类别;词汇这个属性极度稀疏,同样的,衣物标识符这个属性也极为稀疏;最后,词与词之间的相对关系可以通过跳字模型(skip-gram)^[3]或连续词袋模型(continuous bag of words, CBOW)^[3]训练对应词向量进行捕捉,同样的,衣物标识符和合身与否的关系也可以通过训练得到的嵌入矩阵得到。

因此本文首次提出了 SiZeLet (SZL, 恰为本文作者首字母缩写),该模型通过学习用户,衣物与评价等相关稀疏离散属性的嵌入表征,并将其与数值属性相结合,输入进残差网络模块^[4]以预测该衣物对该用户是否合身。

本文的贡献如下:

(1) 本文首次提出了 SiZeLet,该模型在 Product Fit 数据集中验证集上的 F1 分数达到了 0.713 ± 0.01 。

(2) 本文的消融实验证实了用户与衣物的嵌入表征能够有效地提升预测任务的性能表现。

(3) 本文将 fastText 与 GLoVe, textCNN 和 BiRNN 等 NLP 领域技术应用在 SiZeLet 中,并对比了其性能表现。

2 相关工作

Abdulla 等人^[5]采用类似于 word2vec 中的跳字模型,利用购物历史记录以学习衣物的特征,同时该方法通过用户所购买的商品来表征用户,并使用

梯度提升决策树(Gradient Boosting Decision Tree, GBDT)^[6]对合身与否进行预测。Karessli 等人所提出的 SizeNet^[7]将统计模型与从衣物图像所提取的视觉信息相结合,实现了弱监督环境下的衣物合身预测。Dogani^[8]等人不需要具体用户信息,衣物属性或购买反馈的情况下,利用神经协作过滤提出了适用于不同品牌的尺寸预测模型。

3 模型介绍

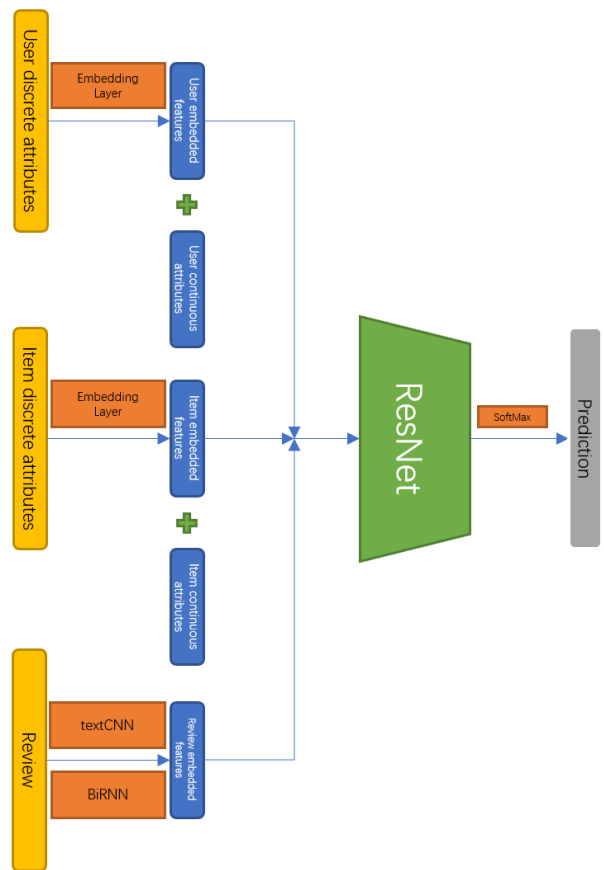


图 1 SiZeLet 模型结构

图 1 为本文提出的 SiZeLet 模型的体系结构,本文将数据的属性分为五类:用户离散属性,例如:身材类型,用户标识符;用户数值属性,例如:身高,体重;衣物离散属性,例如:租借目的,衣物标识符;衣物数值属性,例如:衣物尺寸;文本属性,例如:评价时间,评价摘要,评价详情。本文会利用嵌入层学习得到的嵌入矩阵将用户与衣物的离散属性转换为定长向量并与对应的数值属性进行连结得到相应表征,利用 textCNN^[9]或 BiRNN^[10]载入预训练的词向量以提取文本属性的表征,其中 textCNN 拥有两个嵌入层,其中之一的

权重参与训练，另一个载入预训练的词向量后权重固定，最后将这些表征一同输入进残差网络模块，经过 softmax 运算，最小化交叉熵损失函数对网络进行训练，详细算法可参考算法 1

算法 1. SiZeLet 训练算法

输入：用户离散属性 $X_{user_discrete}$ ；用户数值属性 $X_{user_numeric}$ ；衣物离散属性 $X_{clothe_discrete}$ ；衣物数值属性 $X_{clothe_numeric}$ ；文本属性 X_{review} ；标签 Y ；训练轮次 $epochs$ 。

输出：模型参数 $EmbeddingLayer_{user}$ ； $EmbeddingLayer_{clothe}$ ； $textCNN/BiRNN$ ； $ResNet$ 。

FOR $epoch$ IN $epochs$ DO:

1. $Embedding_{user} = EmbeddingLayer_{user}(X_{user_discrete})$
2. $Embedding_{clothe} = EmbeddingLayer_{clothe}(X_{clothe_discrete})$
3. $Embedding_{review} = textCNN/BiRNN(X_{review})$
4. $Embedding = Embedding_{user} \oplus X_{user_numeric} \oplus$

$Embedding_{clothe} \oplus X_{clothe_numeric} \oplus Embedding_{review}$

5. $L = CrossEntropyLoss(softmax(ResNet(Embedding)), Y)$

6. 利用小批量随机梯度下降法最小化损失函数 L

4 数据集与数据处理

4.1 Product Fit数据集介绍

Product Fit 数据集的训练集包含 135000 条数据，测试集包含 57544 条数据，其中 fit 属性为本课题任务的预测目标，包含 {small, fit, large} 三种类型的标签，本文需要根据每一条数据的其他属性预测该条购买记录中的衣物是否合身。表 1 给出了 Product Fit 数据集的详细信息。

表 1 Product Fit 数据集信息，*表明该属性为预测标签

属性	非空值数量	类别
age	134343	数值属性
body type	124760	离散属性
category	135000	离散属性
weight	114160	数值属性
height	134513	数值属性
item_id	135000	离散属性
rating	134944	数值属性
bust size	122115	数值属性
rented for	134995	离散属性
review_date	135000	离散属性
review_summary	134766	离散属性
review_text	134954	离散属性
size	135000	数值属性

user_id	135000	离散属性
fit*	135000	离散属性

4.2 数据预处理

4.2.1 数值属性预处理

Product Fit 数据集中的某些属性虽然为数值属性，但是存储时却以字符串形式存储，例如：weight, height 和 bust size，因此需要将其转换为数值属性。

(1) weight 属性在本数据集中的表达形式为“数字 lbs”，因此本文使用相应的正则表达式对其中的数字进行捕捉，从而将其转换为数值属性。

(2) height 属性在本数据集中的表达形式为“英尺’英寸””，因此本文使用相应的正则表达式对其中的英尺与英寸进行捕捉，并将其转换为以厘米为单位的数值属性。

(3) bust size 属性无疑是本数据集中数值属性中最为特殊的一个，因为其完全可以被视为离散属性，但是本文发现可以通过（下胸围，上下胸围差）这样一个二维向量即可以表示不同胸围之间的相对关系，因此本文通过表 2 对胸围属性进行了映射，将其转换为二维向量。

表 2 胸围属性映射关系

罩杯	数值
aa	0.5
b	1
c	2
d	3
d+	4
dd	5
ddd/e	6
f	7
g	8
h	9
i	10
j	11

4.2.2 离散属性预处理

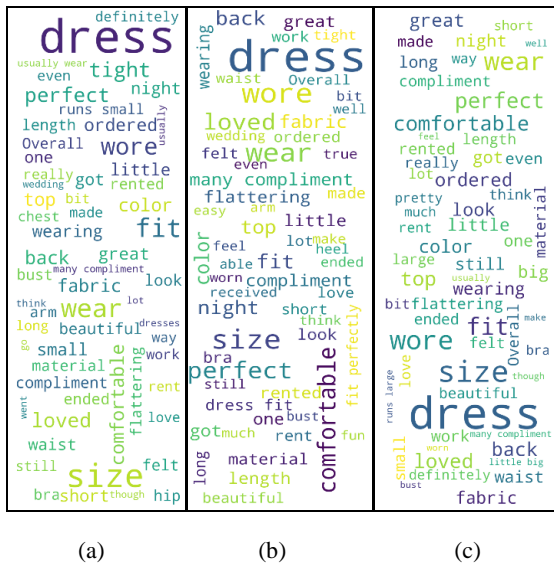


图2 (a) (b) (c) 分别为 small, fit, large 标签下的 review_text 所生成的词云

(1) 评价属性: 本文认为衣物的合身与否与购买时间并无强相关性, 因此将 review_date 属性直接删除。

图2展示了 small, fit, large 标签下的 review_text 所生成的相应的词云, 可以看出, 除了 dress, size, fit 这些与衣物本身有着强相关性的词语出现在了所有标签下的词云之外, small 标签下 review_text 所生成的词云则出现了 small 和 tight 等这些表明尺寸较小的形容词; fit 标签下 review_text 所生成的词云则出现了 perfect, flattering 和 compliment 等这些带有积极意味的褒义评价词; 而 large 标签下 review_text 所生成的词云则出现了 long, large 和 big 等这些表明尺寸较大的形容词, 同时其词云中出现的 little 也经常与 big 共同出现。并且每个标签所对应的词云中都有着只在自身词云中出现的特征词, 因此衣物合身与否很大程度上就能通过评价所展现, 因此有效地提取文本属性中的特征能够极大地提高本课题任务的性能。

本文首先将 review_summary 与 review_text 拼接为 review 属性, 其次将其中出现次数少于一定次数的词语删去, 并将其填充至固定长度, 最后通过在不同语料集上不同方式下预训练好的词向量将其映射为固定长度的向量, 其中参数设置可以参考本文 5.1 抑或本文的开源代码。

(2) 用户及衣物的离散属性: Product Fit 数据集中出现的其他离散属性, 例如 category, item_id 和 user_id 等, 均会通过 SiZeLet 中的相应维度为(属

性种类数, 嵌入维度数) 嵌入矩阵映射为固定长度的嵌入向量, 这也是本文标题《万物皆可嵌入》的直接体现, 其中嵌入层的参数也是通过训练中学习得到的。表3展示了相应离散属性的种类数目。

表3 Product Fit 数据集中离散属性的种类数目

属性	种类数目
body_type	7
rented_for	9
category	68
item_id	5850
user_id	105571

4.2.3 缺失数据处理

从表1可以看出 Product Fit 数据集中的某些属性存在大量的缺失值, 因此本文使用了 scikit-learn^[11]对相关缺失值进行了处理, 其中数值属性会使用对应属性下的中位数进行填充, 离散属性会使用对应属性下出现频次最高的元素进行填充。

4.2 数据标准化

如果输入的数值属性具有非常大的比例差异, 往往会导致深度神经网络模型表现不佳, 因此本文对所有的数值属性进行了标准化处理。

5 实验

5.1 实验设置

实验中的相关设置可以参考表4, 本表中的所有设置均可以在本文的开源代码的 config.py 中进行修改与组合, 欢迎读者自行探索。

表4 SiZeLet 模型设置

实验设置	数值
用户/衣物离散属性	10
嵌入维度	
评价属性嵌入维度	100
textCNN 卷积核大小	[3, 4, 5]
textCNN 通道数大小	[100, 100, 100]
BiRNN 隐藏层个数	2
BiRNN 隐藏层维度	100
用户数值属性维度	5
衣物数值属性维度	2
激活函数	ReLU
学习率	0.001

预训练词向量方式	GLoVE
预训练词向量语料库	twitter.27B
训练集/验证集	0.8/0.2
最小词频	300
填充长度	300

性嵌入表征	
GLoVE	0.713±0.01
fastText	0.692±0.008
textCNN	0.713±0.01
BiRNN	0.651±0.02

5.2 实验结果

相关实验结果可以参考表 5。在表 4 实验配置下的 SiZeLet 在训练集上的 F1 分数取得了 0.713±0.01 的成绩。

表 5 SiZeLet 在 Product Fit 验证集上的性能表现

Accuracy	F1 Score	AUC
0.831±0.005	0.713±0.01	0.886±0.002

5.3 消融实验

本文所提出的 SiZeLet 的另一个优势在于其较为优秀的低耦合性，用户可以选择是否使用用户/衣物离散属性嵌入表征，也可以选择使用 fastText^[12]或是 GloVe^[13]，也可以选择使用 textCNN 或 BiRNN，这也为后续的消融实验提供了可能。

表 6 展示了相关消融实验下的结果，除了相应配置的改变，其他配置均与表 4 保持一致。从表 6 中可以看出：

(1) 用户/衣物离散属性嵌入表征的使用使得 F1 分数提升了 2.5%，这证实了其有效性。

(2) GloVe 相对于 fastText 在本课题任务上更具有优势。

(1) textCNN 相对于 BiRNN 的准确率更高，值得一提的是，在 PyTorch1.6, RTX2080Ti 的配置下，使用 textCNN 的 SiZeLet 每一轮 epoch 训练的时间平均为 43.3s，而使用 BiRNN 的 SiZeLet 则是 65.6s，因此 textCNN 在性能表现上相对于 BiRNN 也更具有优势。

综上，本文所推荐的设置是使用用户/衣物离散属性嵌入表征 + GloVe + textCNN。

表 6 消融实验结果	
配置	F1 Score
使用用户/衣物离散属性嵌入表征	0.713±0.01
不使用用户/衣物离散属性嵌入表征	0.688±0.01

6 思考与展望

6.1 思考

在真实的线上购买场景中，电商平台需要在交易发生前，只根据用户与衣物的属性为用户预测该衣物的尺码是否合适，因为评价通常是在交易结束后产生，所以此时是没有评价属性参与预测的，所以本文为了模拟真实的线上购买场景，只使用了用户与衣物的属性参与训练，最终 SiZeLet 的 F1 分数为 0.432±0.005，虽然性能表现远不及使用评价属性的 SiZeLet，但是该种假设下的 SiZeLet 更具有普适意义。

6.2 展望

当前的 SiZeLet 仍然存在着一些不足，例如：当新增了用户或者新增了衣物，当前的网络则需要重新训练，以学习新的用户的衣物的嵌入表征，因此在 SiZeLet 中实现增量学习是一个可行的解决方案。

同时，因为嵌入表征之间的关系可以从余弦相似度反映，例如：在 GloVe 中，beautiful 与 lovely 的余弦相似度为 0.921，因此我们也可以对衣物的嵌入向量进行聚类，将其聚为 3 类，分别对应 {small, fit, large}，随后电商平台可以对 small 与 large 类中的衣物进行检查以发现衣物是否在设计或生产流程中存在问题，这个问题有待在未来的 SiZeLet 中得到解决。

致谢 感谢 CoSec 实验室所提供的计算资源：RTX2080Ti 与 RTX3090；感谢 The Stanford Natural Language Processing Group 所开源的预训练词向量；感谢李宇峰老师与助教们所提供的 Product Fit 数据集。

参 考 文 献

- [1] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.
- [2] Köppen M. The curse of dimensionality[C]//5th Online World Conference on Soft Computing in Industrial Applications (WSC5). 2000, 1: 4-8.
- [3] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[J]. arXiv preprint arXiv:1310.4546, 2013.
- [4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [5] Abdulla G M, Borar S. Size recommendation system for fashion e-commerce[C]//KDD Workshop on Machine Learning Meets Fashion. 2017.
- [6] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. Annals of statistics, 2001: 1189-1232.
- [7] Karessli N, Guigourès R, Shirvany R. Sizenet: Weakly supervised learning of visual size and fit in fashion images[C]//Proceedings of the IEEE/CVFP Conference on Computer Vision and Pattern Recognition Workshops. 2019: 0-0.
- [8] Dogani K, Tomassetti M, Vargas S, et al. Learning Embeddings for Product Size Recommendations[C]//eCOM@ SIGIR. 2019.
- [9] Zhang Y, Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1510.03820, 2015.
- [10] Maas A, Daly R E, Pham P T, et al. Learning word vectors for sentiment analysis[C]//Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies. 2011: 142-150.
- [11] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python[J]. the Journal of machine Learning research, 2011, 12: 2825-2830.
- [12] Bojanowski P, Grave E, Joulin A, et al. Enriching word vectors with subword information[J]. Transactions of the Association for Computational Linguistics, 2017, 5: 135-146.
- [13] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543.

Background

A Fitting clothing can not only reduce the return cost of users due to size problems, but also improve user satisfaction. Therefore, suggesting suitable sizes for users based on their body shapes is important for e-commerce platforms. This paper proposes SiZeLet for the first time. The network extracts embedding representations of relevant discrete attribute values from the three dimensions of user, clothing, and review, and combines it with continuous attribute values to predict whether

clothing fits the user. The F1 score of the model on the validation set of the Product Fit dataset has reached 0.713 ± 0.01 , and the effectiveness of the embedding representation was confirmed through ablation experiments. At the same time, this article uses this as an opportunity to compare fastText with GLoVE, textCNN and BiRNN and other NLP technologies. Finally, the relevant code of this article will be open sourced at <https://github.com/ZhiliShen/SiZeLet>.