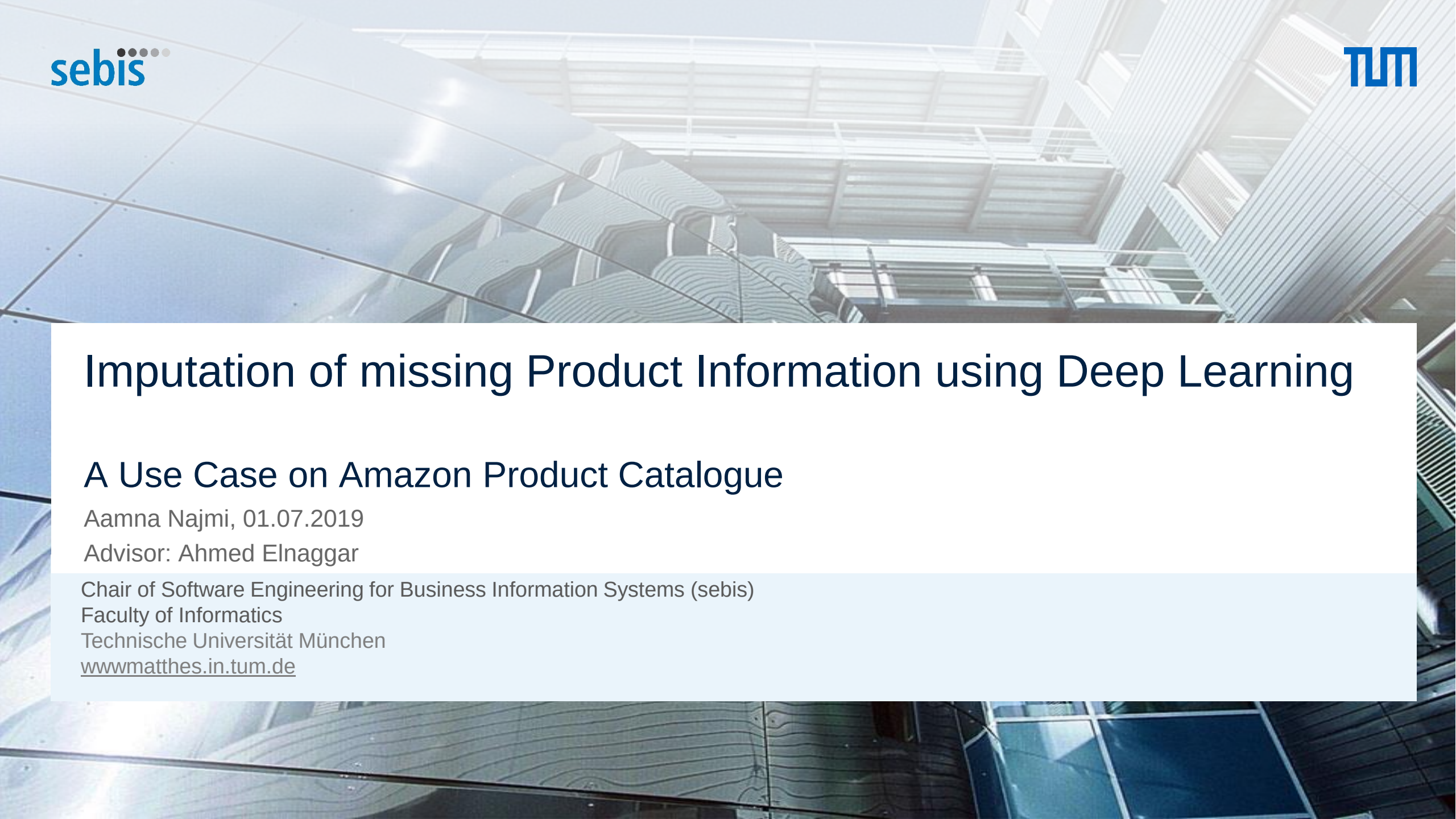


Imputation of missing Product Information using Deep Learning

A Use Case on Amazon Product Catalogue

Aamna Najmi, 01.07.2019

Advisor: Ahmed Elnaggar

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
wwwmatthes.in.tum.de

Outline

1. Motivation
2. Research Questions
3. Approach
4. Implementation
5. Evaluation
6. Conclusion

© sebis

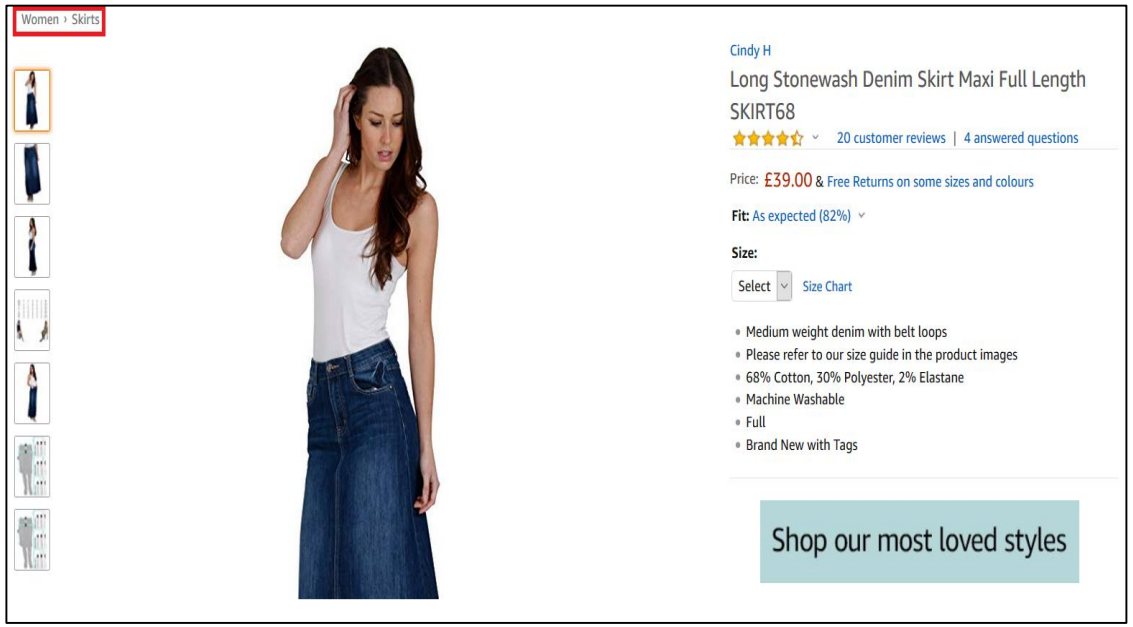
Outline

1. Motivation
2. Research Questions
3. Approach
4. Implementation
5. Evaluation
6. Conclusion

© sebis

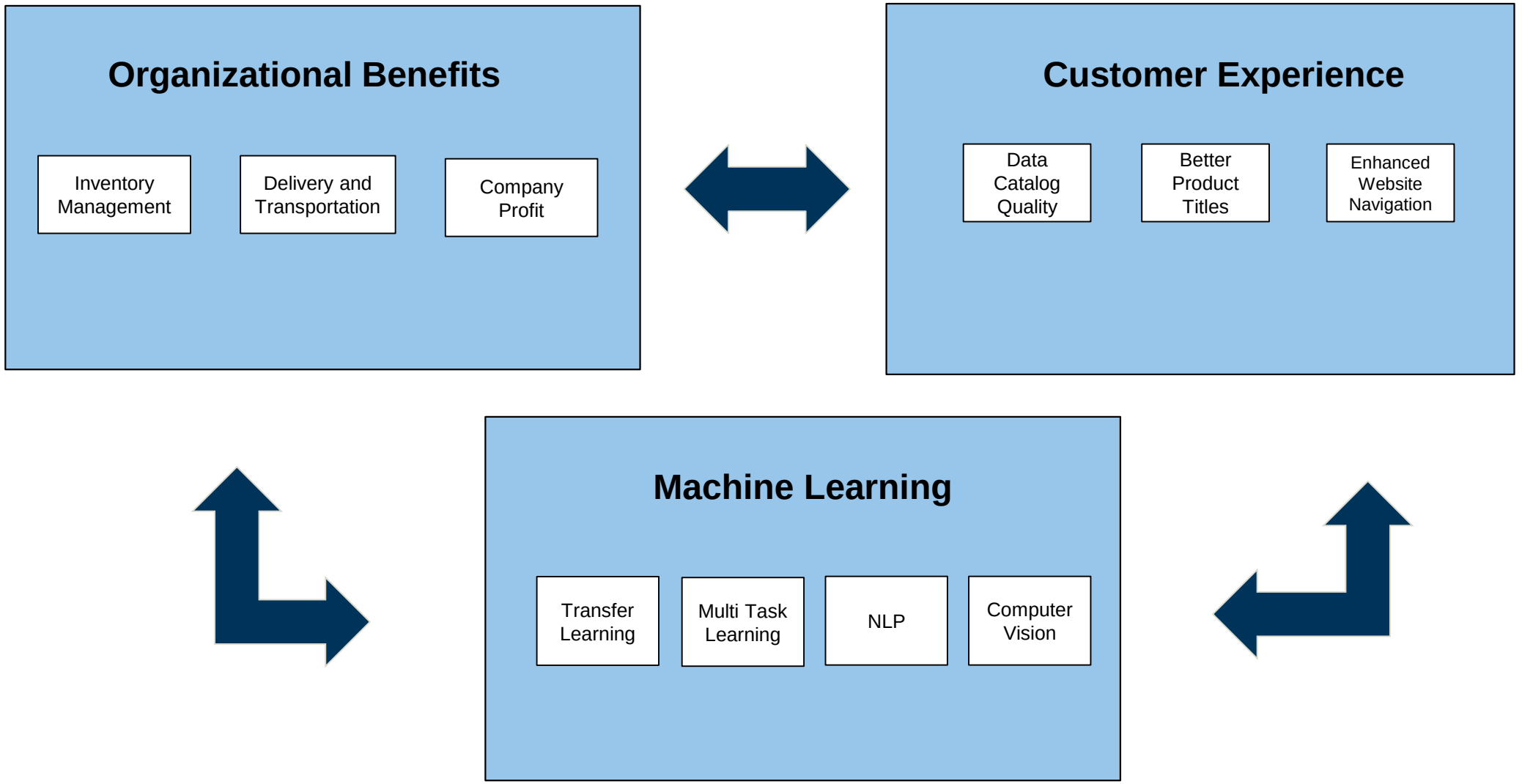
Motivation

- Global retail ecommerce sales will reach about $4 trillion in 2020, accounting for 14.6% of total retail spending worldwide [1]
- 20% of purchase failures are potentially a result of missing or unclear product information [2]
- Detailed product information = improved customer experience and company profit



© sebis

Motivation



© sebis

Outline

1. Motivation
2. Research Questions
3. Approach
4. Implementation
5. Evaluation
6. Conclusion

© sebis

Research Questions

1	Could Multi-task Learning (MTL) and Transfer Learning (TL) perform better than Single Task Learning (STL) on the Amazon Product Catalog dataset?
2	What architecture choices and hyperparameters shall we use in both Multi-task and Transfer Learning to obtain good performance?
3	Can Transfer Learning and Multi-task Learning be useful in the e-commerce domain to enhance user-experience?

© sebis

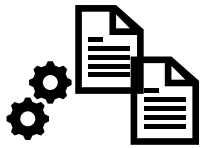
Outline

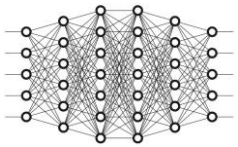
1. Motivation
2. Research Questions
3. Approach
4. Implementation
5. Evaluation
6. Conclusion

© sebis

Approach

Methodology





Web Scraping → Preparing the dataset → Integrate dataset → Train the model on STL, MTL and TL architectures → Evaluation of the results → Verify and validate the research questions

© sebis

Approach

Single Task Learning

- Each task is trained independently
- Model is trained in isolation and is task specific
- Network approximates a function for output of a single task only
- Same model cannot be good in generalizing for other tasks

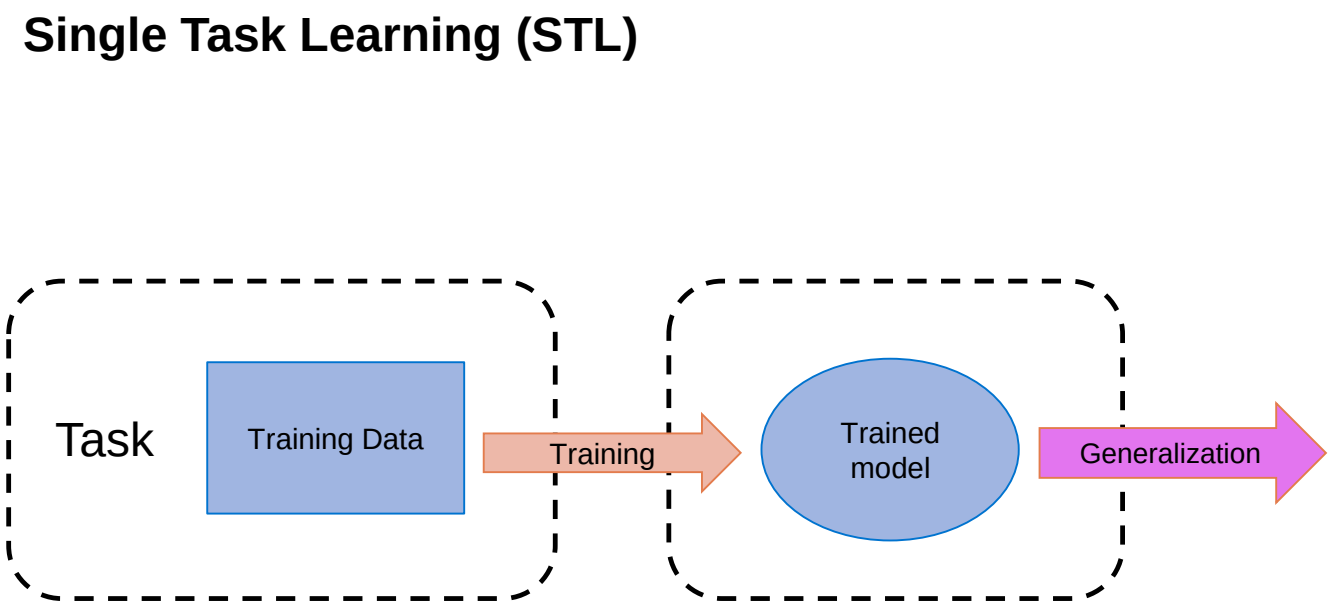


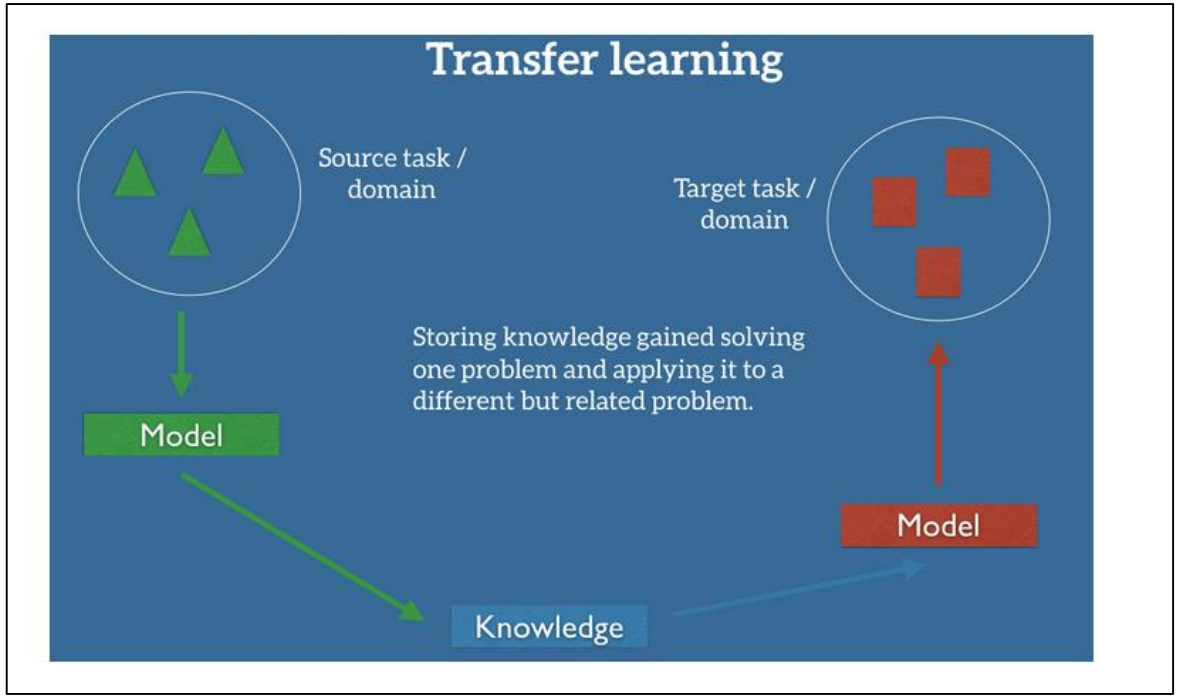
Image courtesy: Multi-Task Learning: Theory, Algorithms, and Applications SDM 2012

© sebis

Approach

Transfer Learning

- Domain adaptation technique
- Source and target domain have different datasets and the tasks may or may not be same [3]
- Pre-trained networks trained on a large dataset like Imagenet [4]
- Source dataset generally much larger in size
- Overcome issues like class imbalance problem and shortage of data



© sebis

Approach

Multi-task Learning

- Domain adaptation technique
- Promotes generalization over multiple tasks [5]
- Tasks trained in parallel
- Optimizes more than one loss function
- Form of inductive transfer
- Can be of two variants: Soft parameter or Hard parameter sharing [6]

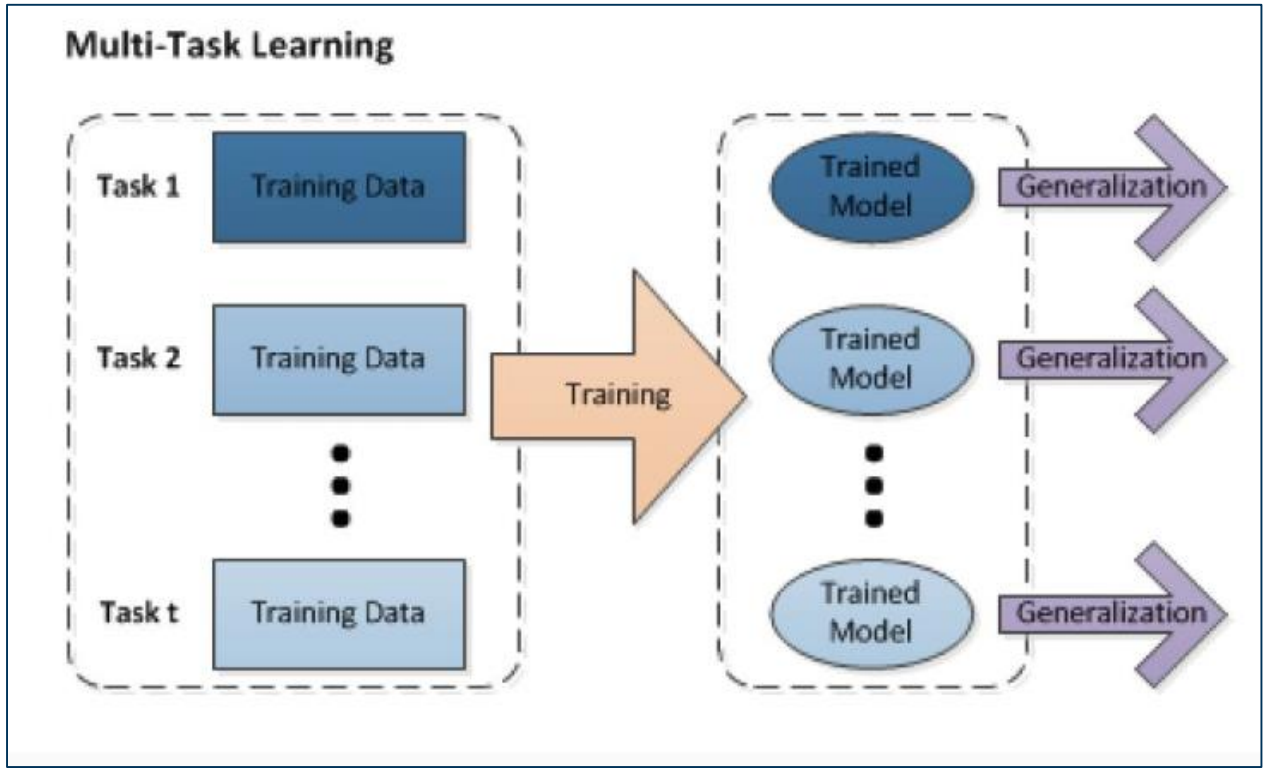


Image courtesy: Multi-Task Learning: Theory, Algorithms, and Applications
SDM 2012

© sebis

Outline

1. Motivation
2. Research Questions
3. Approach
4. Implementation
5. Evaluation
6. Conclusion

© sebis

Implementation

Dataset Overview

- **Source**: Scraped data from five regional Amazon websites (UK, DE, FR for text based and UK, DE, IT for image based)
- **Records**: ~200k from each website
- **Attributes**: Product ID, Product Title, Product Description, Color, Category, Brand, Target Gender, Product Summary, Product Specifications, Product Image

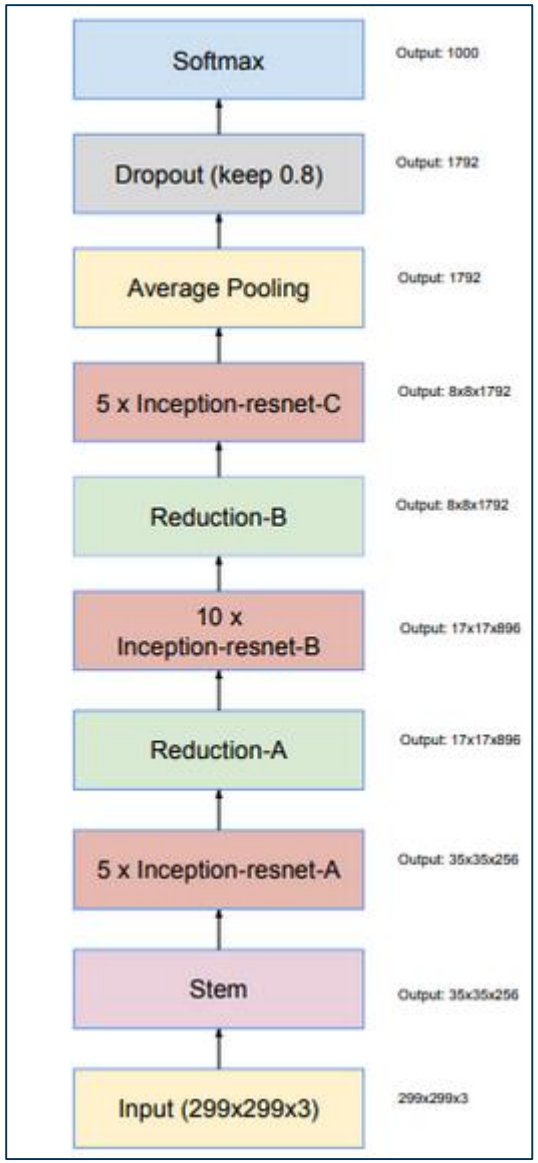
- **Samples**

- *Nike* Slam *Women's* Dri-Fit Tennis *Skirt* - *Black*-XL: Amazon.co.uk: Clothing.
- *Aerolite* Leichtgewicht 4 Rollen *Trolley* Koffer Kofferset Gepäck-Set Reisekoffer Rollkoffer Gepäck, 3 teilig, *Schwarz/Grau*: Amazon.de: Koffer, Rucksäcke & Taschen.
- *Diesel* pour *Homme* Waykee L.32 *Pantalon* - *Noir* - 32 W/34 L: Amazon.fr: Vetements et accessoires.

Implementation

Image based models

- State of the art Convolutional Neural Network (CNN) called Inception Resnet V2 [7]
- The network model has outperformed previous state of the art architectures on the Imagenet dataset challenge that involves image classification task.
- Use the architecture with variations in neural blocks and training schemes for our tasks.



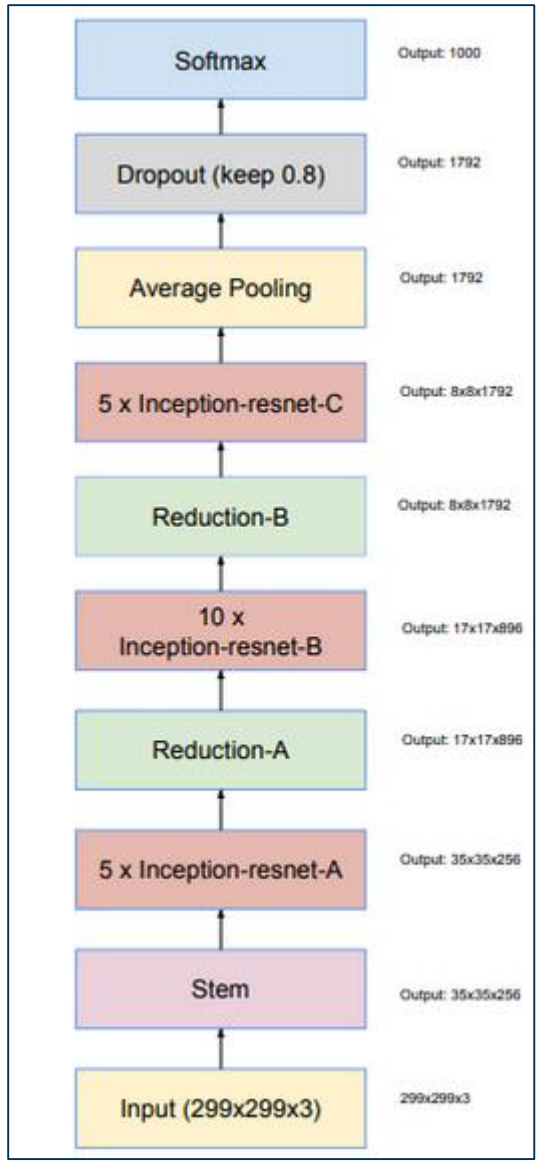
© sebis

Implementation

Image based models

Single Task Learning

- Trained Inception Resnet Model V2 from scratch
- Integrated the image dataset for each of the 12 tasks with the model independently
- Cross Entropy as the Loss function
- Stochastic Gradient Descent with momentum as optimizer



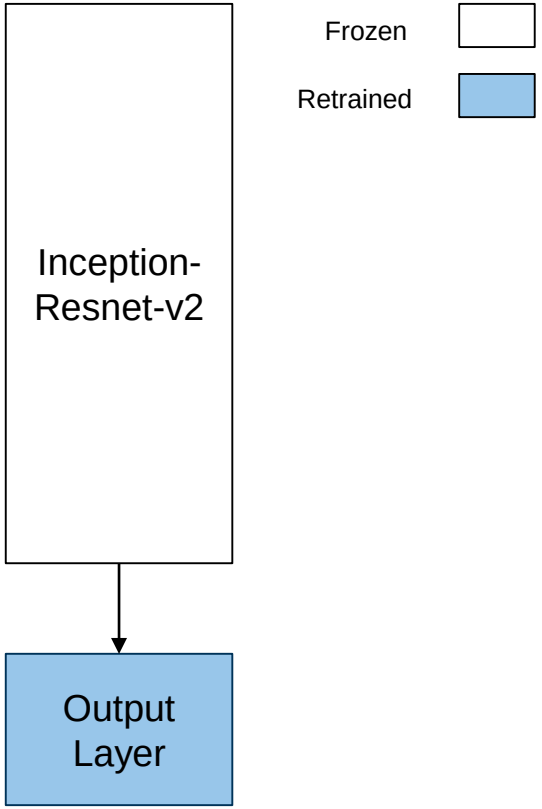
© sebis

Implementation

Image based models

Transfer Learning- Strategy I

- Used pretrained version of the Inception Resnet V2 Model
- Model trained on the Imagenet benchmark dataset
- This strategy involves freezing the pretrained layers and adding an additional output layer at the end
- Only the Output layer is retrained
- Cross Entropy as the Loss function
- Stochastic Gradient Descent with momentum as optimizer



The first version of the network used to implement Transfer Learning on images

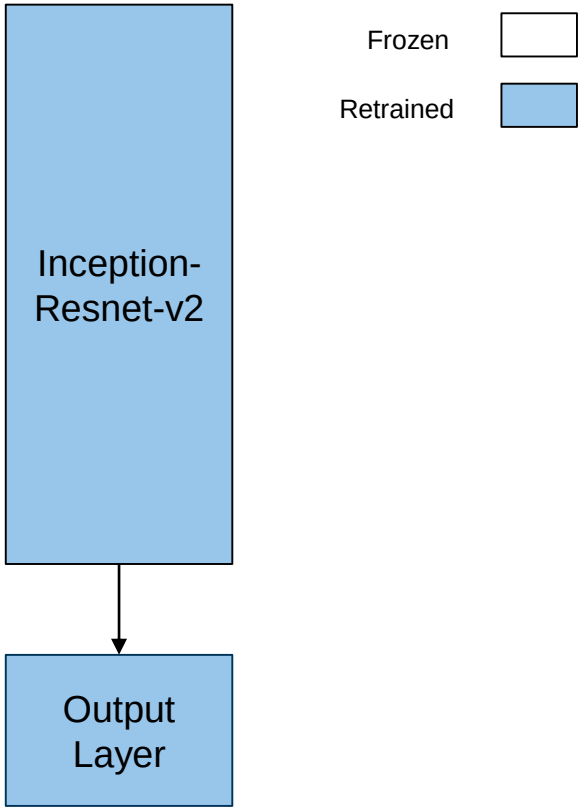
© sebis

Implementation

Image based models

Transfer Learning- Strategy II

- Used pretrained version of the Inception Resnet V2 Model
- Model trained on the Imagenet benchmark dataset
- This strategy involves involves initializing the network with the pretrained weights adding an additional output layer
- The entire network alongwith the Output layer is retrained
- Cross Entropy as the Loss function
- Stochastic Gradient Descent with momentum as optimizer



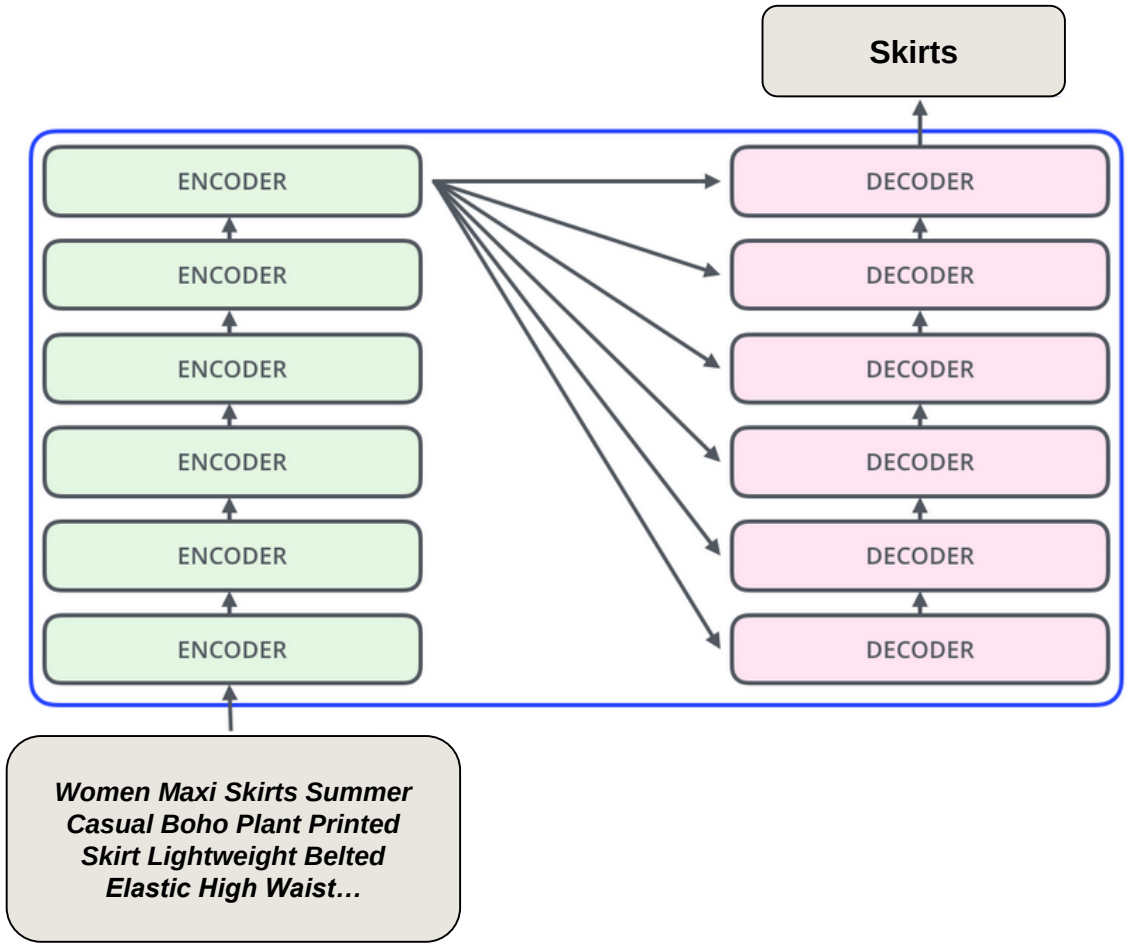
The second version of the network used to implement Transfer Learning on images

© sebis

Implementation

Text based models

- State-the-art sequence to sequence model called **Transformer**
- Uses an attention mechanism that works well with text related tasks [8]
- Consists of encoder and decoder unit with attention mechanisms [9]
- Each layer of encoder and decoder consists of a feed forward layer
- The transformer model is implemented and trained using the Tensor2Tensor (T2T) library



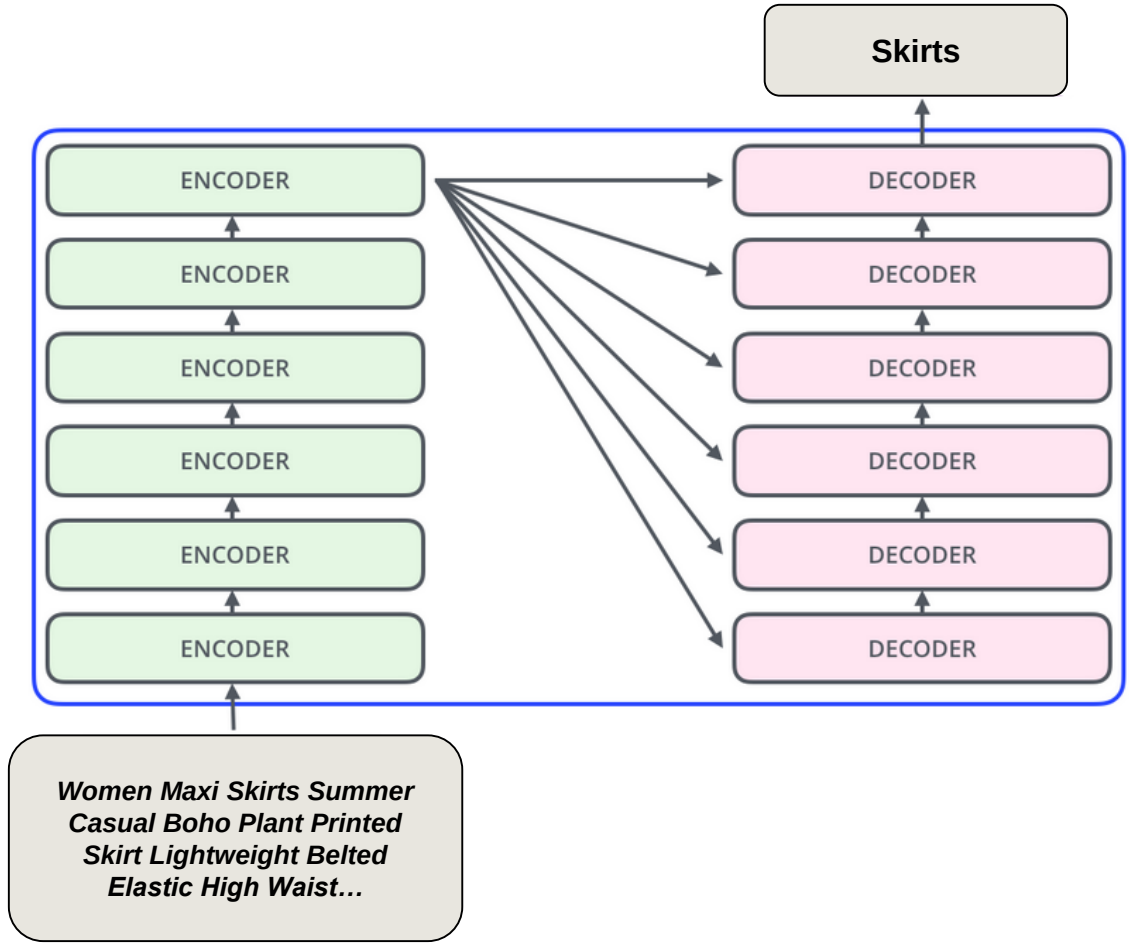
© sebis

Implementation

Text based models

Single Task Learning

- Used the Transformer model
- Trained a separate model for every text based task
- 12 independent models for 12 tasks
- Implemented single task training using T2T



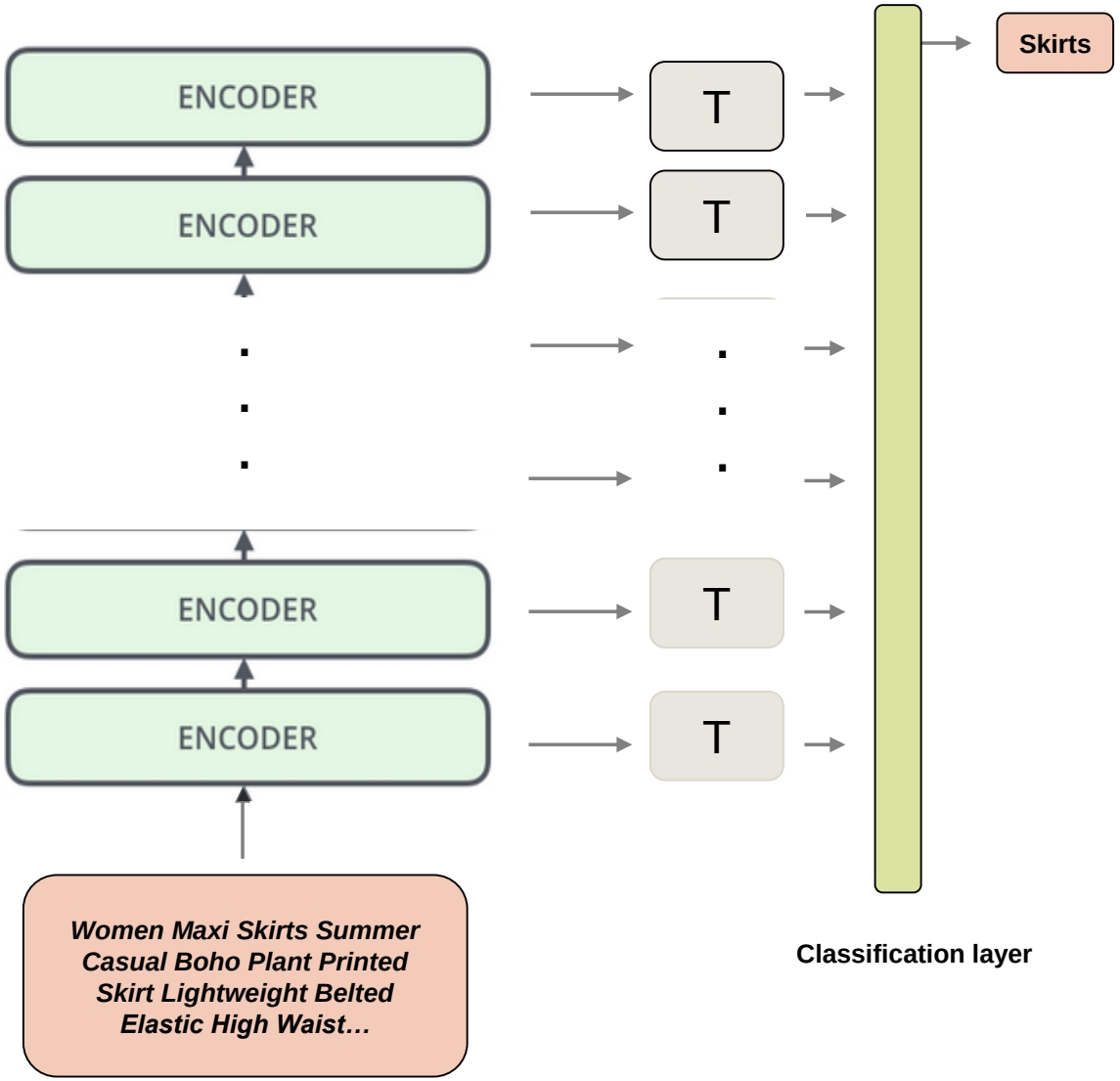
© sebis

Implementation

Text based models

Transfer Learning

- Context aware network called BERT (*Bidirectional Encoder Representations from Transformers*) [10]
- The model is pretrained on a language modeling task using the Wikipedia corpus in French, German and English
- Used the base version of BERT which consists of 12 layers, 12 self attention heads and a hidden size of 768 [11]
- The pretrained model weights are frozen and only the last layer for classification is used get the class scores

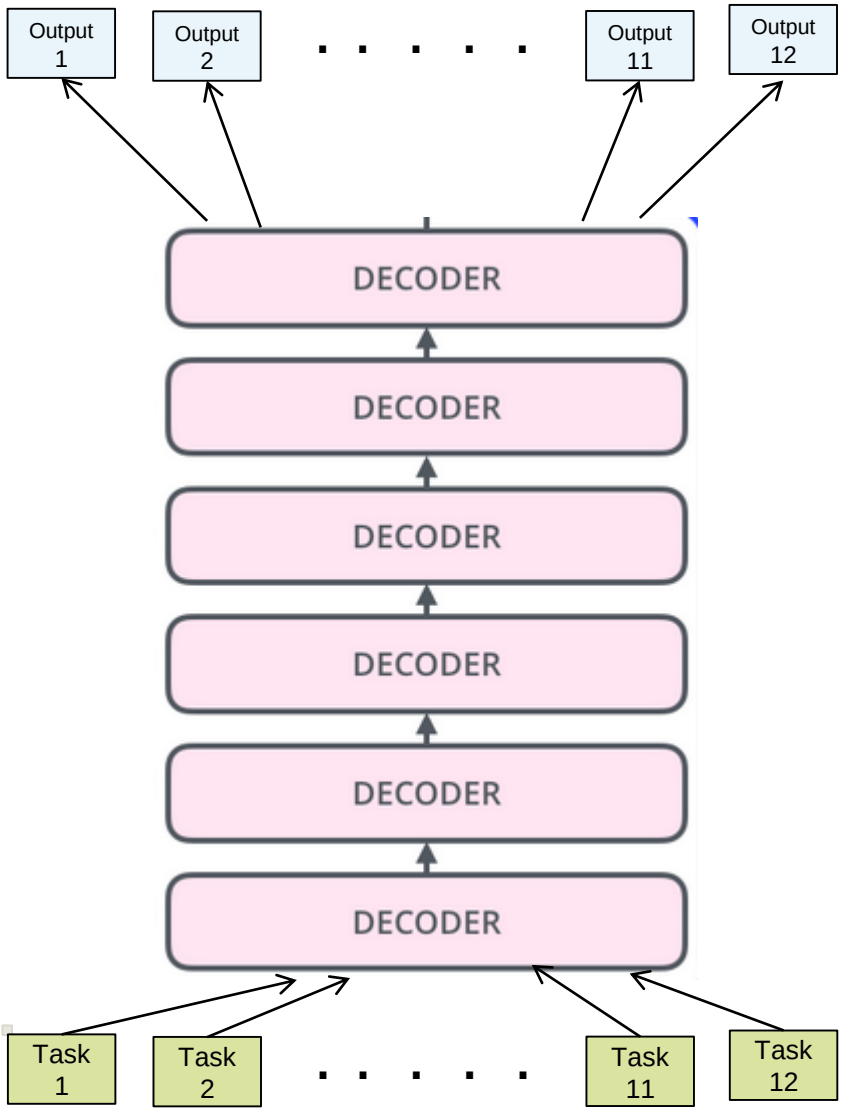


Implementation

Text based models

Multi-task Learning

- Trained all the 12 tasks across languages and task families concurrently using the Transformer
- Additional Language Model (LM) task based on Wikipedia corpus consisting of English, French, Romanian and German tokens.
- The 12 text based tasks are appended after the LM task
- All the tasks use the LM vocabulary to generate and integrate the data using T2T
- Multiple loss functions optimized concurrently



© sebis

Outline

1. Motivation
2. Research Questions
3. Approach
4. Implementation
5. Evaluation
6. Conclusion

© sebis

Evaluation

Experimental Setup

- Hardware

	Machine
Name	DGX-1
GPUs	8x Tesla V100
Core	41k
Memory	8x 16GB

- Software

Programming Language	Python
Scraping	Requests Method, Beautiful Soup package
Text based models	Tensor2Tensor framework, tensorflow
Image based models	Pytorch framework, pretrainedmodels package

© sebis

Evaluation

Experimental Setup (contd.)

- Hyperparameters

Text based

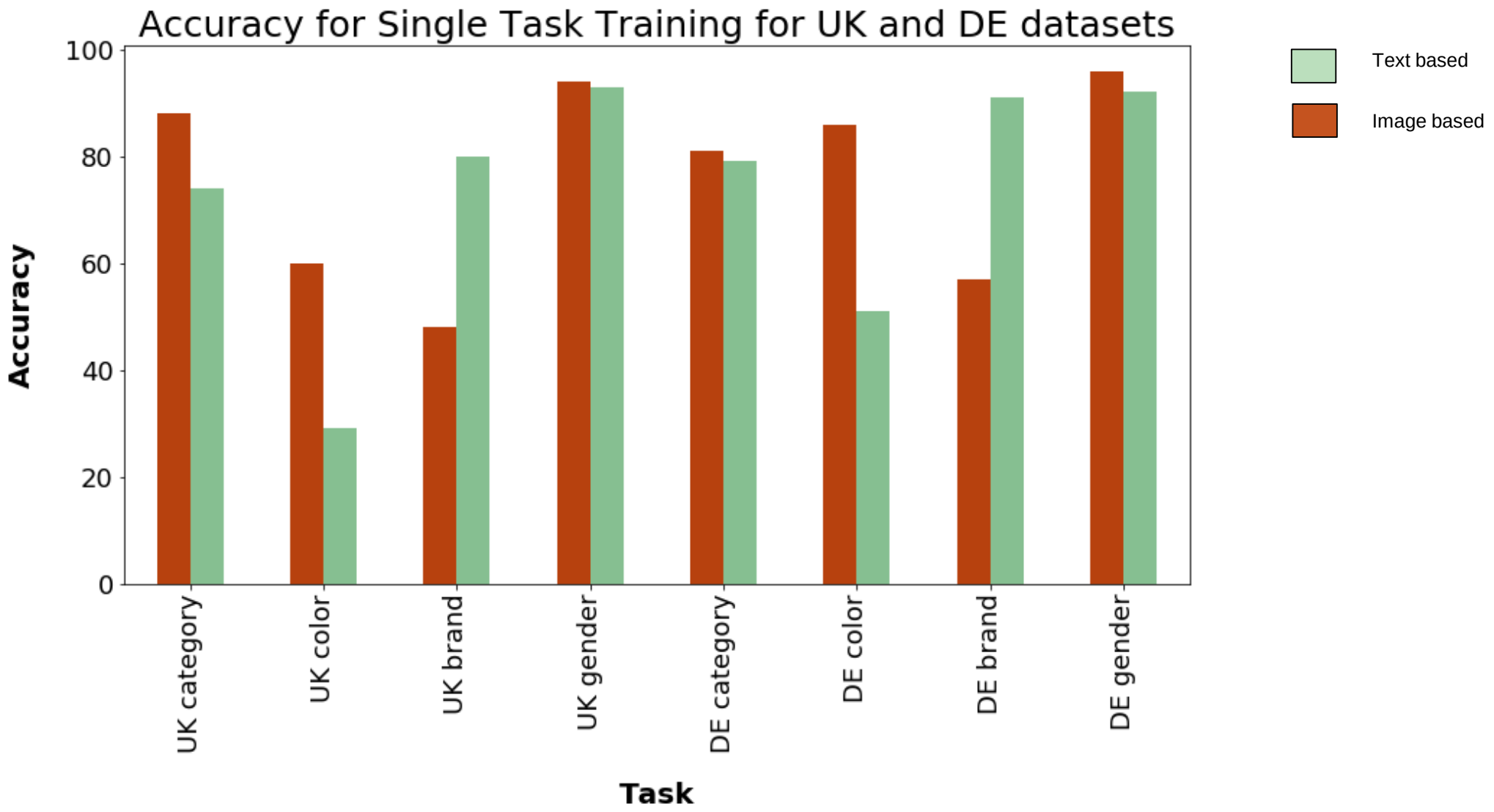
	Single Task	Transfer Learning	Multi-task Learning
Model	Transformer	BERT	Transformer
Hidden size	512	768	1024
Filter size	2048	3072	8192
Batch Size	4096	32	1024
Optimizer	Adam	Adam	Adam
Maximum sequence length	300	300	512
GPUs used	1	1	8

Image based

	Single Task	Transfer Learning: Strategy I	Transfer Learning: Strategy II
Model	Inception-Resnet-V2	Inception-Resnet-V2	Inception-Resnet-V2
Fine tuning	No	Last Layer only	Whole Network
Pre-trained	No	Yes	Yes
Batch size	196	1024	196
Optimizer	Adam	Adam	Adam
GPUs used	7	1	7

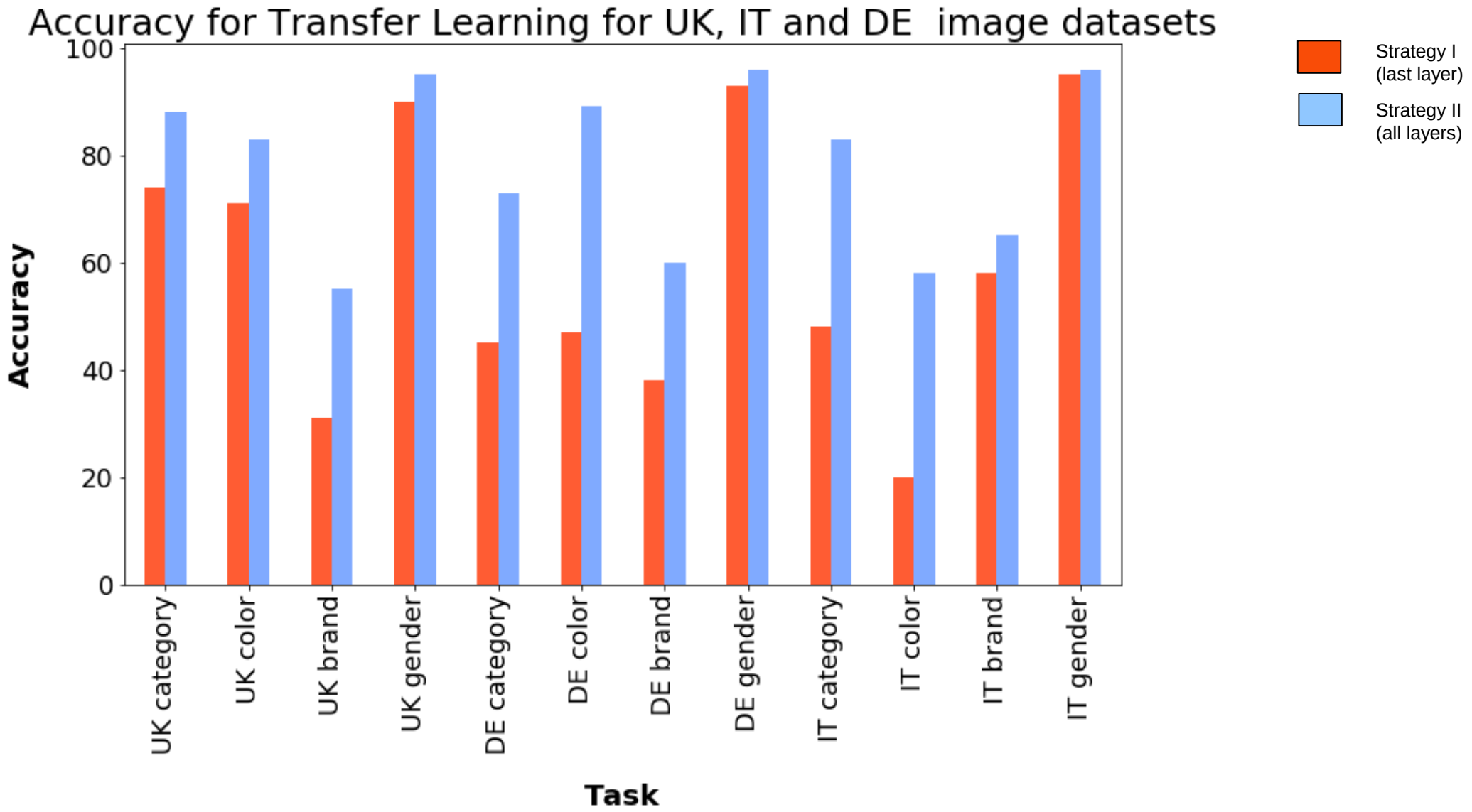
© sebis

Evaluation- Single Task Training



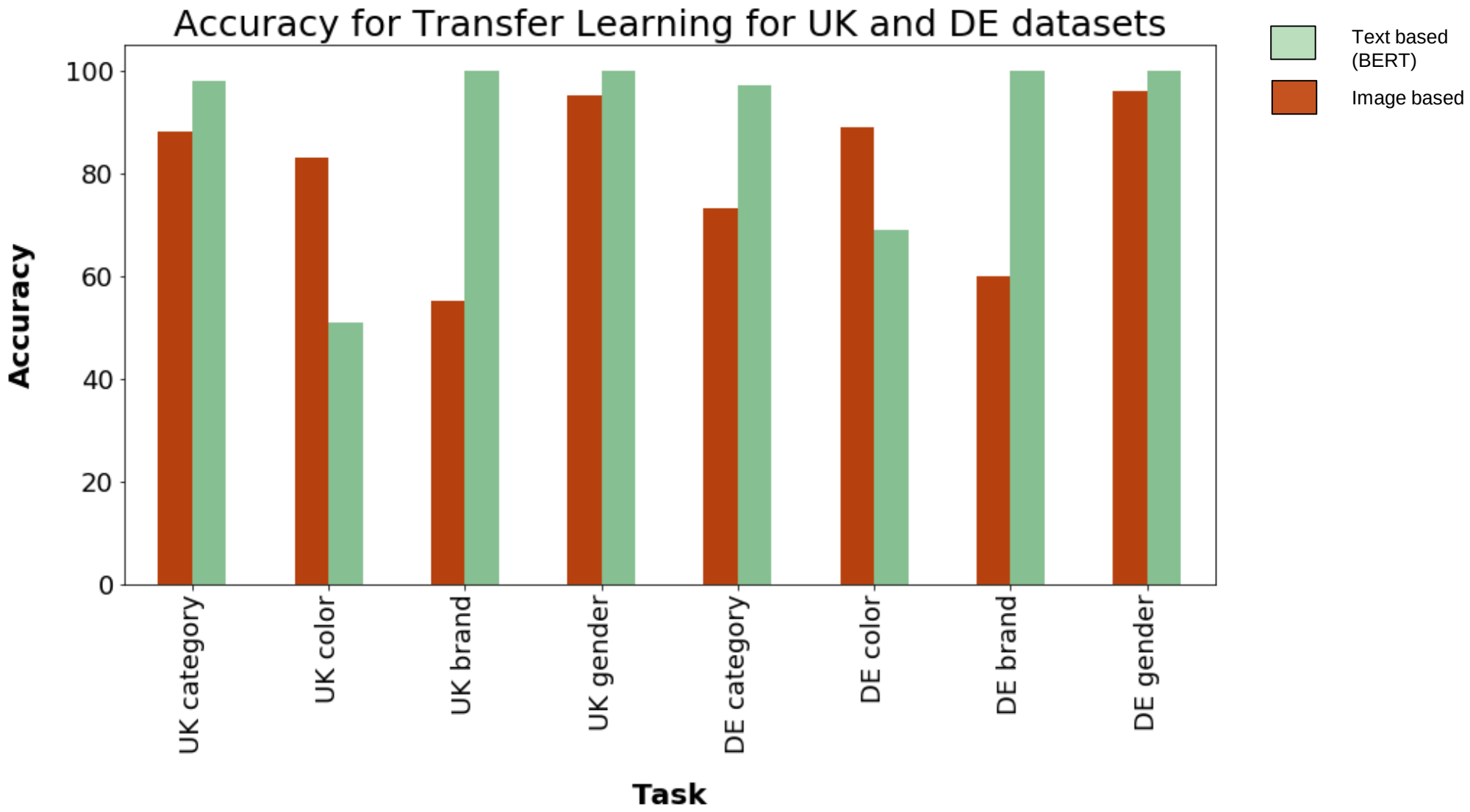
© sebis

Evaluation- Transfer Learning



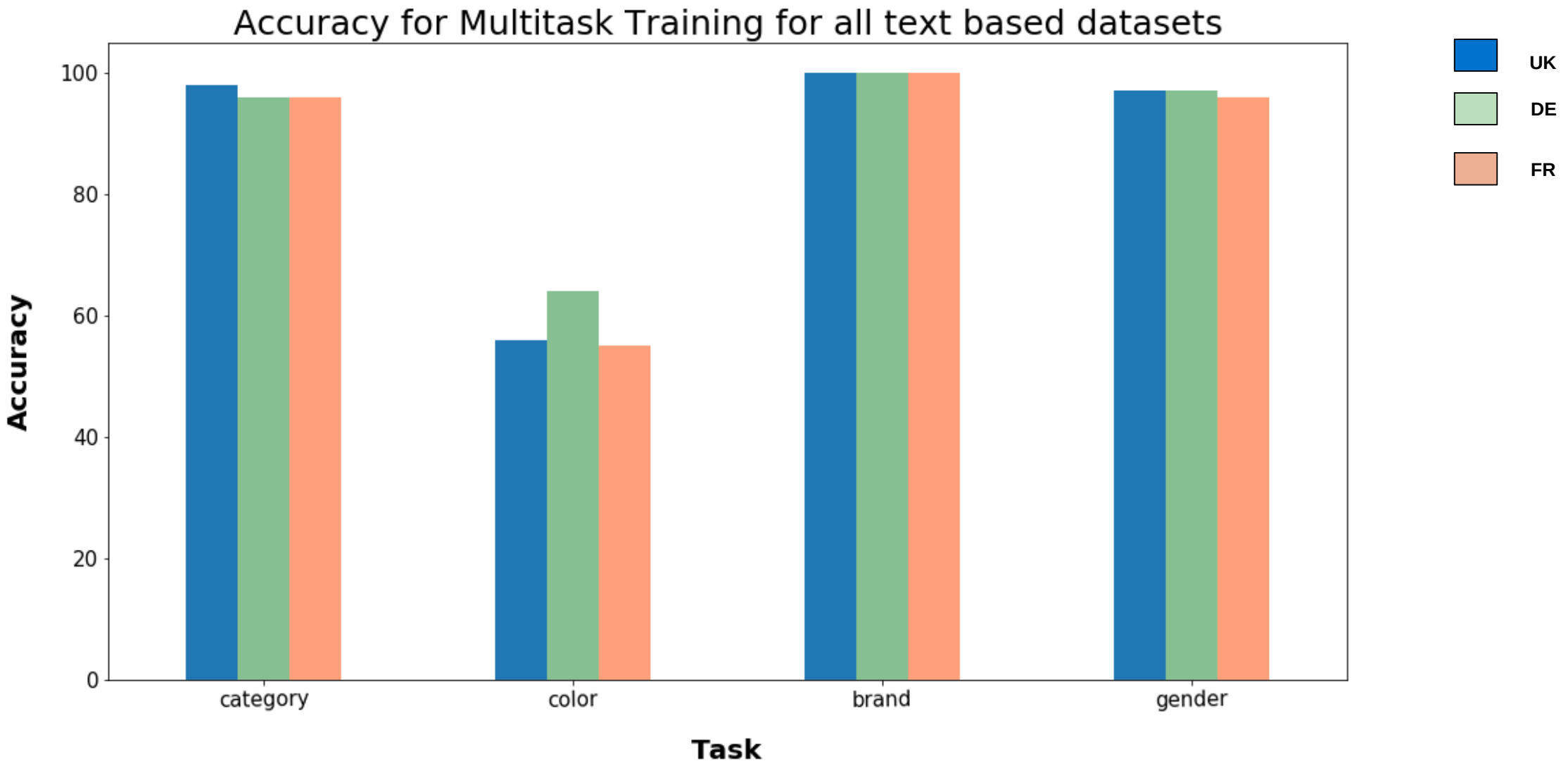
© sebis

Evaluation- Transfer Learning



© sebis

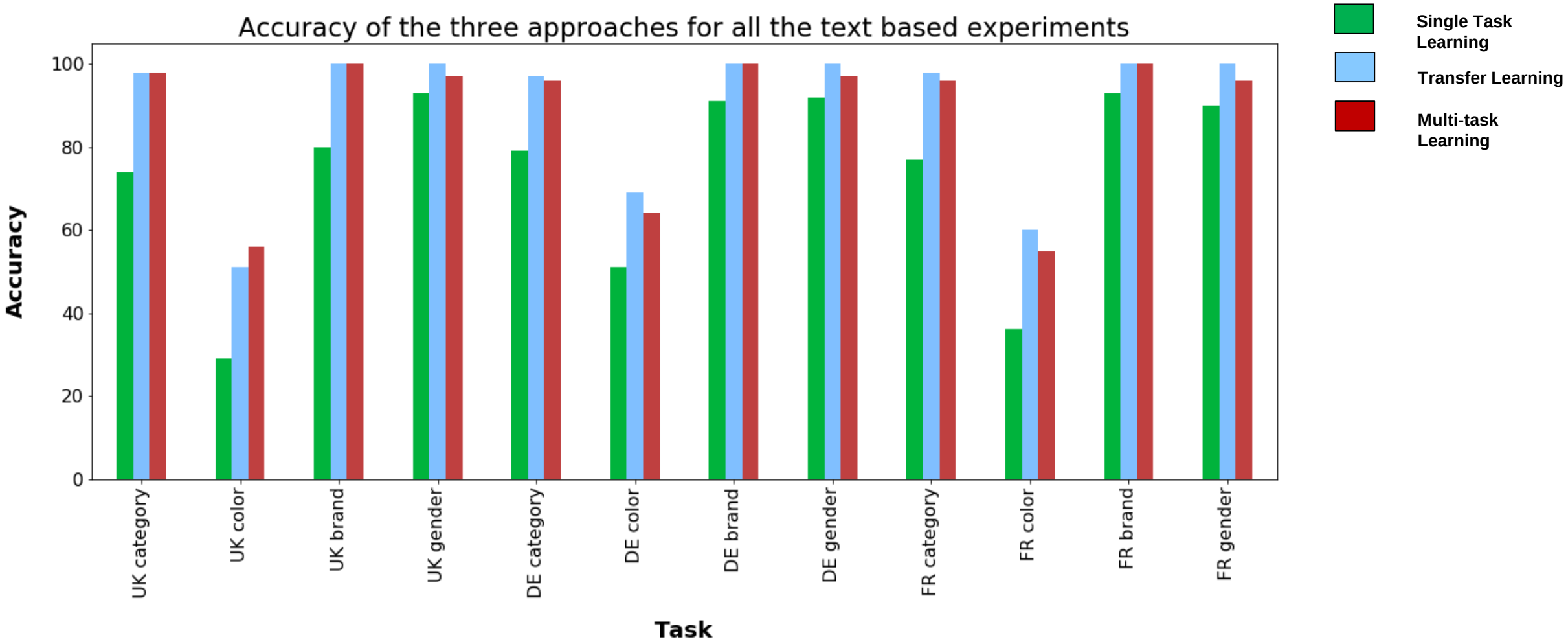
Evaluation- Multi-task Learning



© sebis

Evaluation- Across Task Approaches

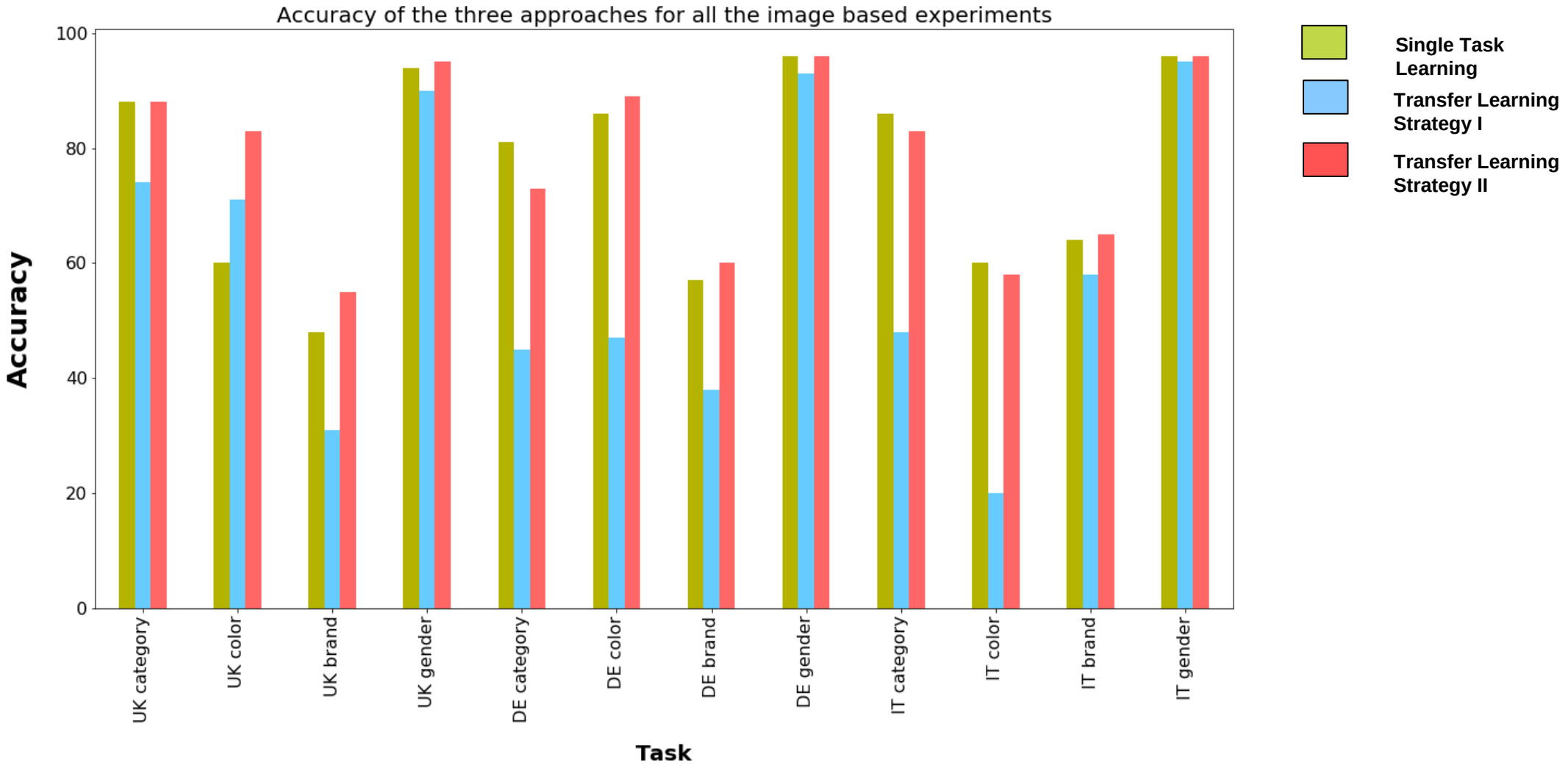
- Text based



© sebis

Evaluation- Across Task Approaches

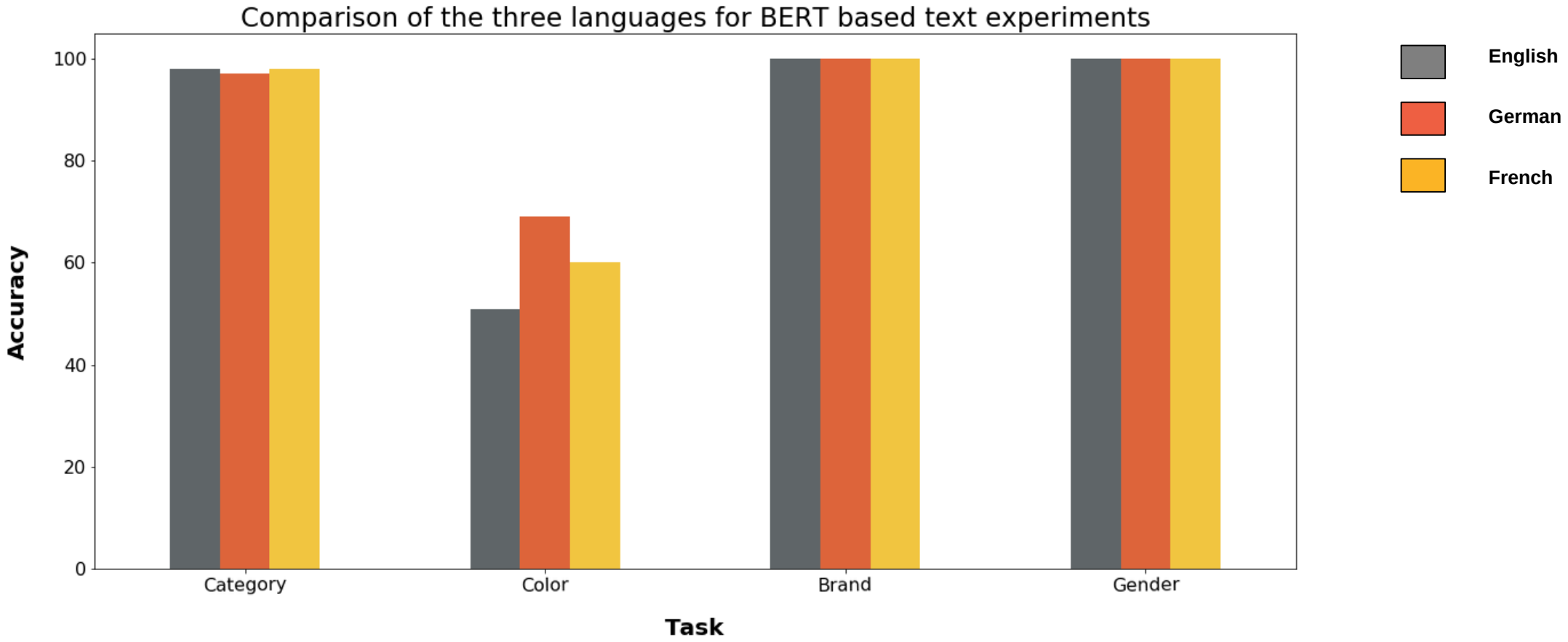
- Image based: STL requires 2x more time to train in comparison with TL



© sebis

Evaluation

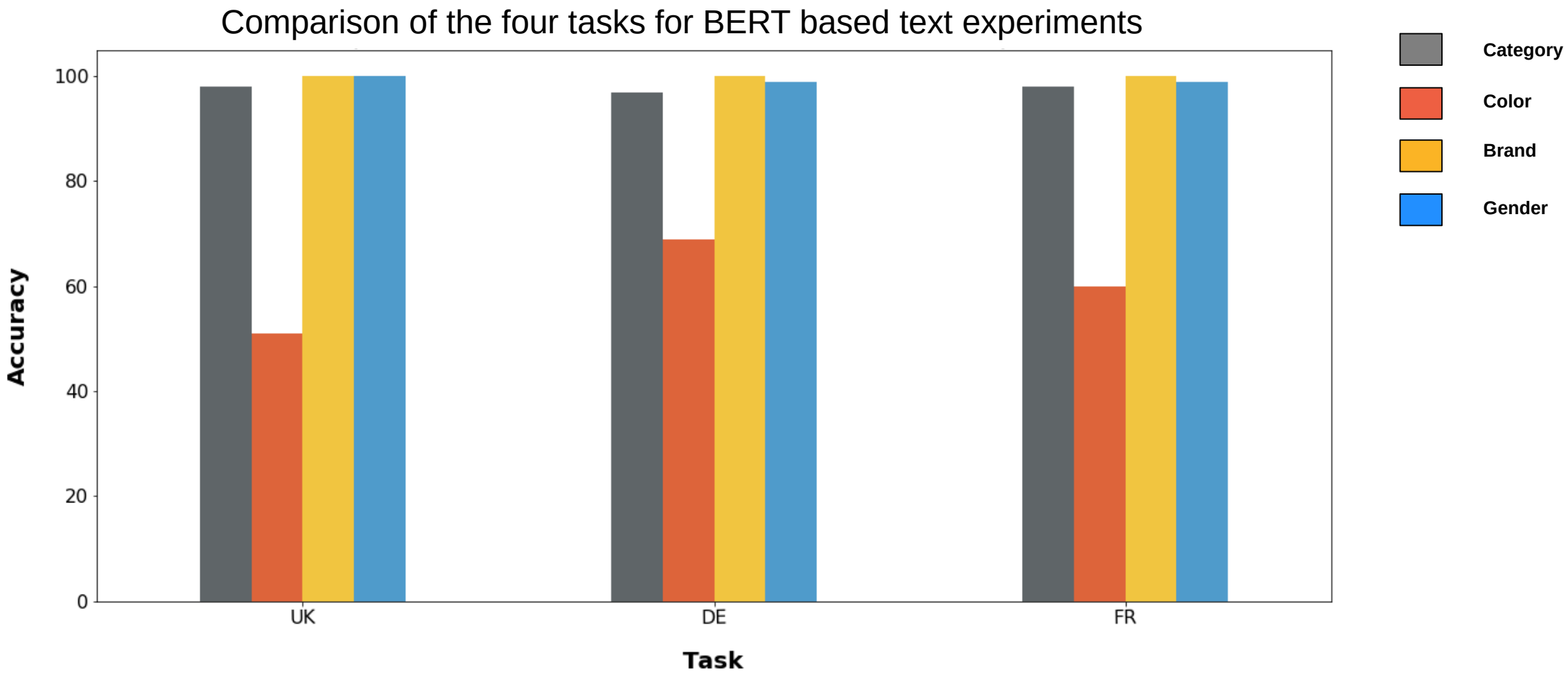
Across Task Languages



© sebis

Evaluation- Across Task Families

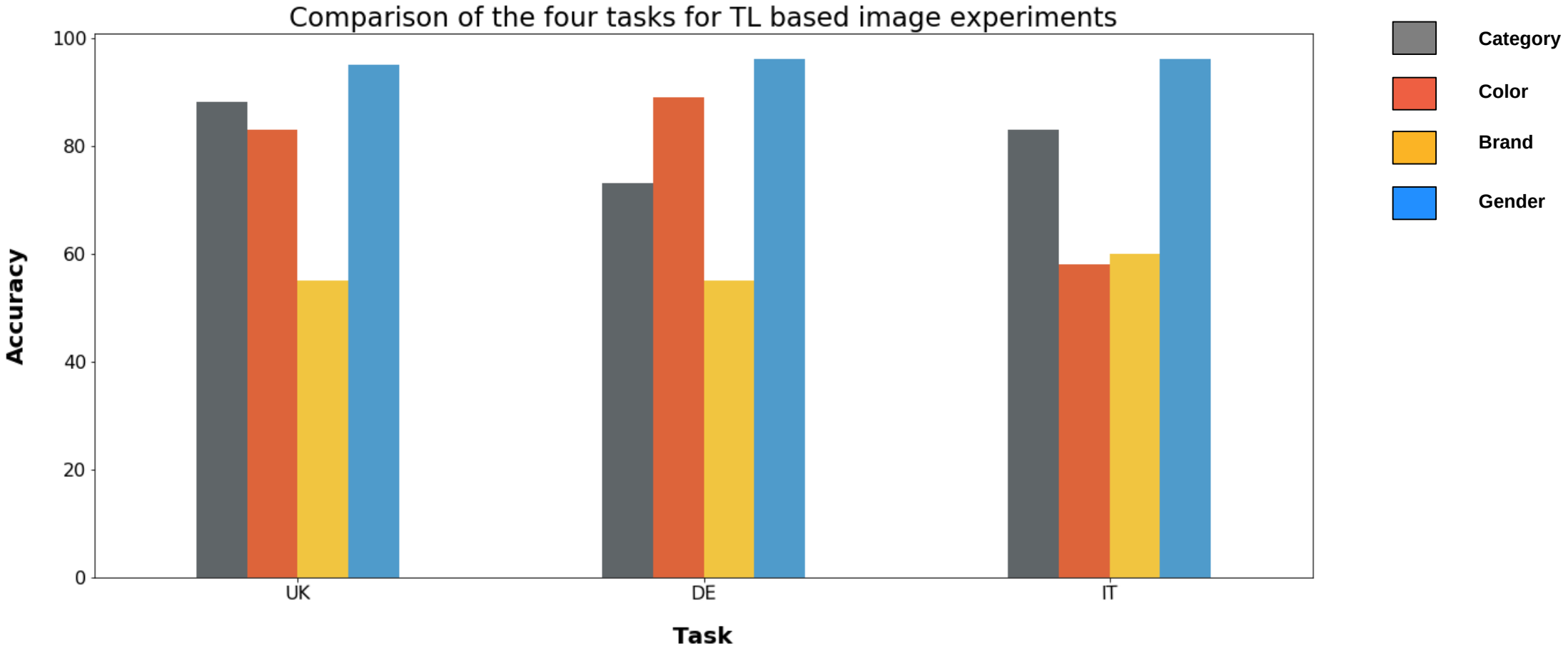
Text based



© sebis

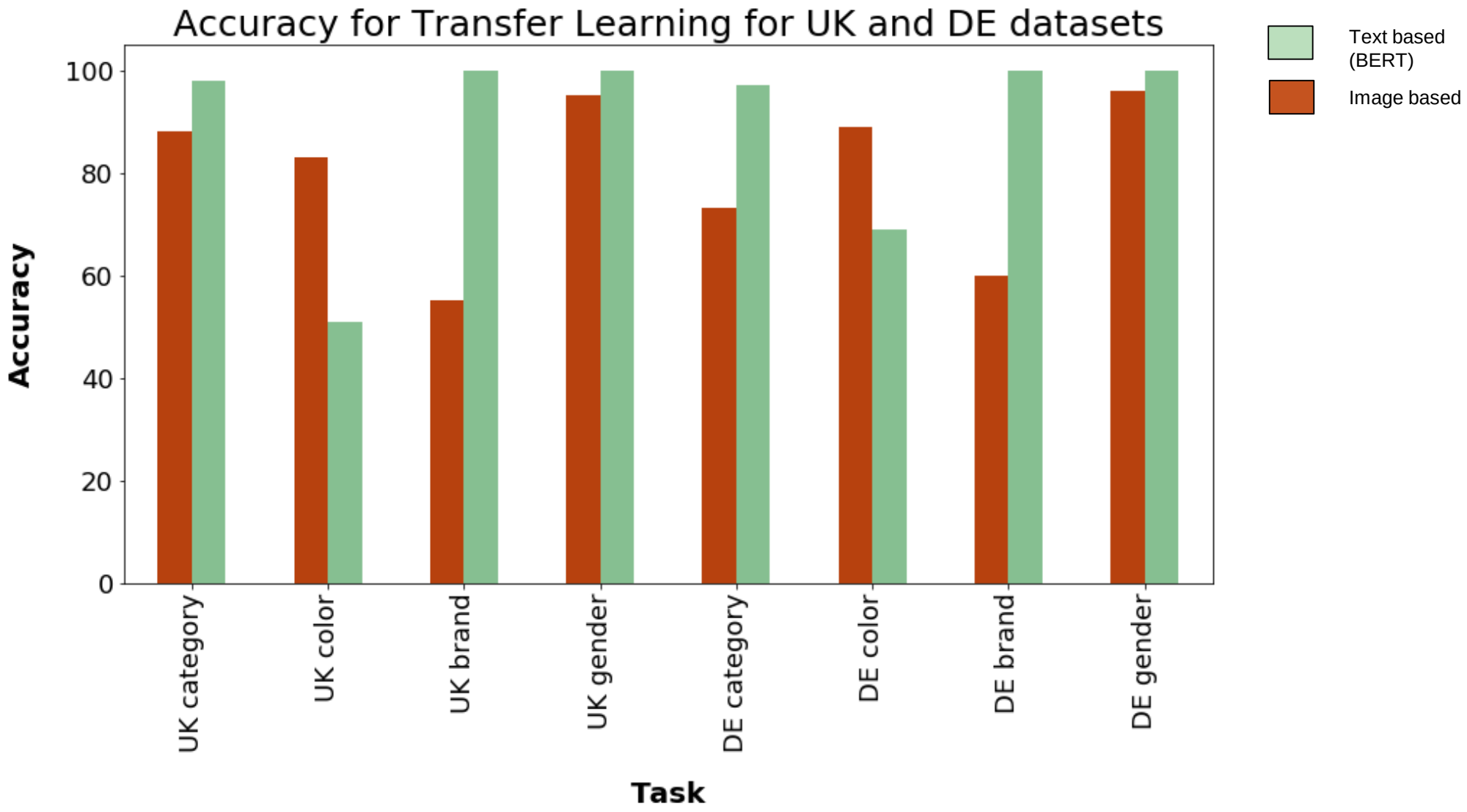
Evaluation- Across Task Families

Image based



© sebis

Evaluation- Across Task Modalities



© sebis

Evaluation- Survey

12 questions

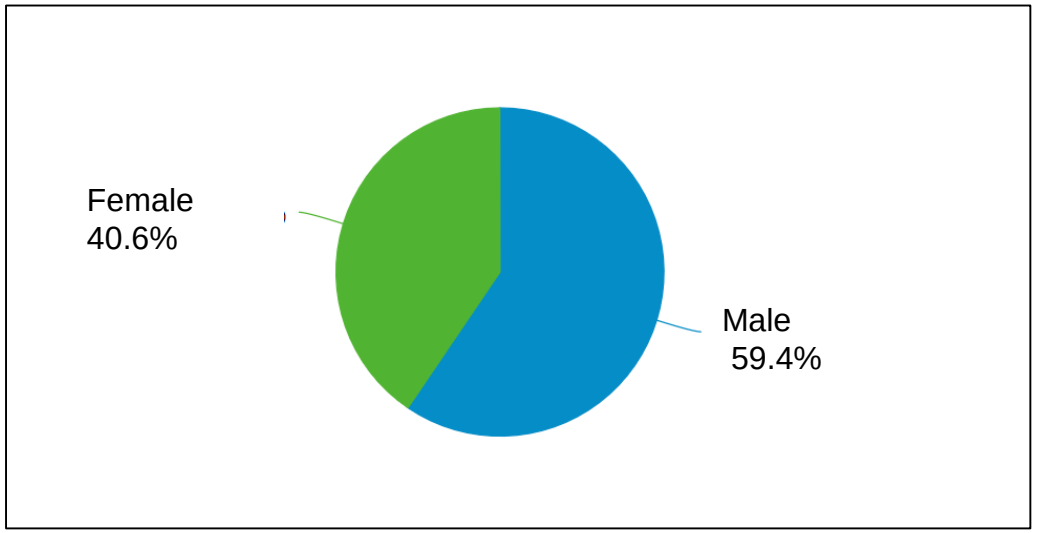
79 respondents

32

23

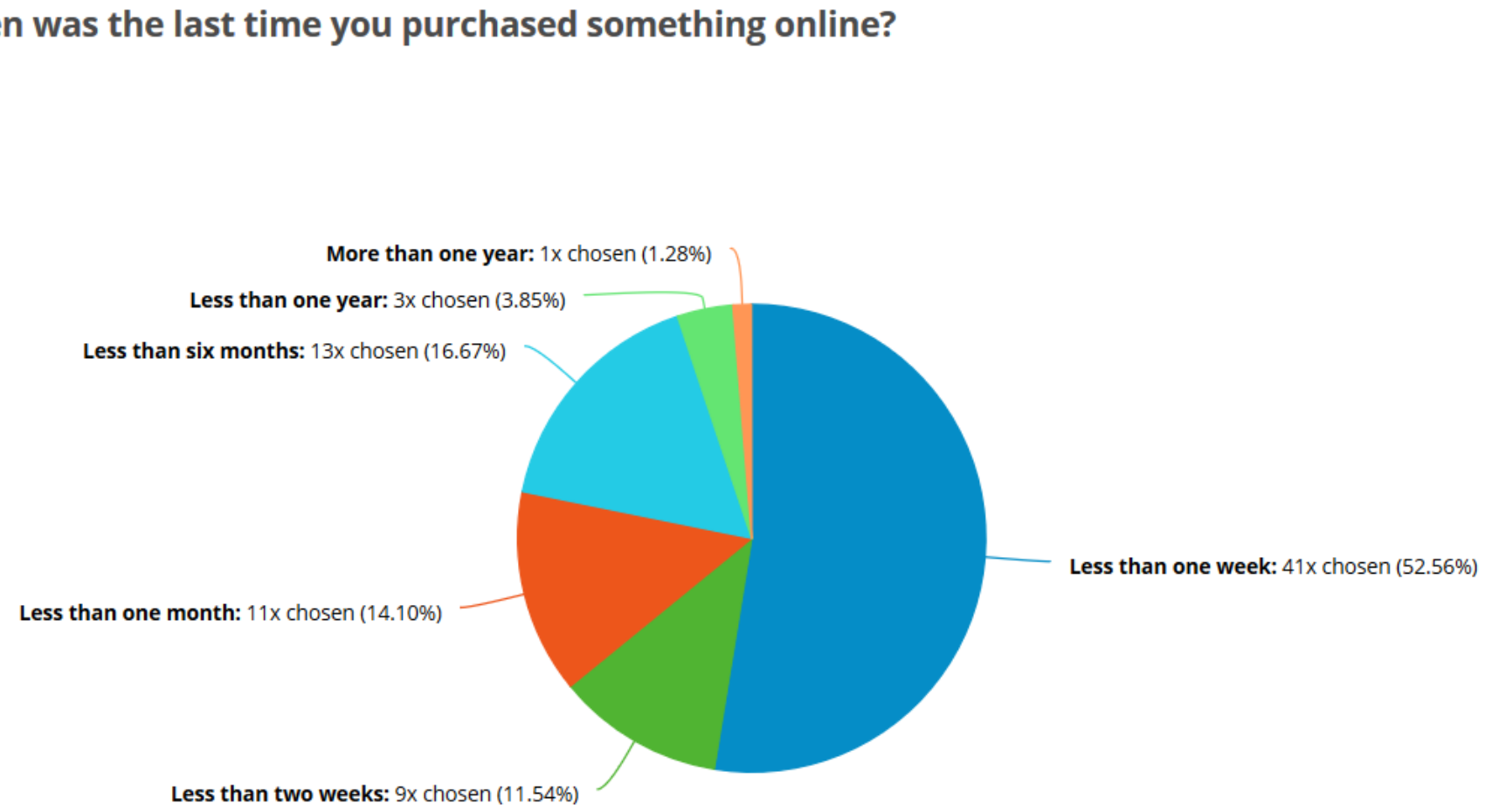
12

7



© sebis

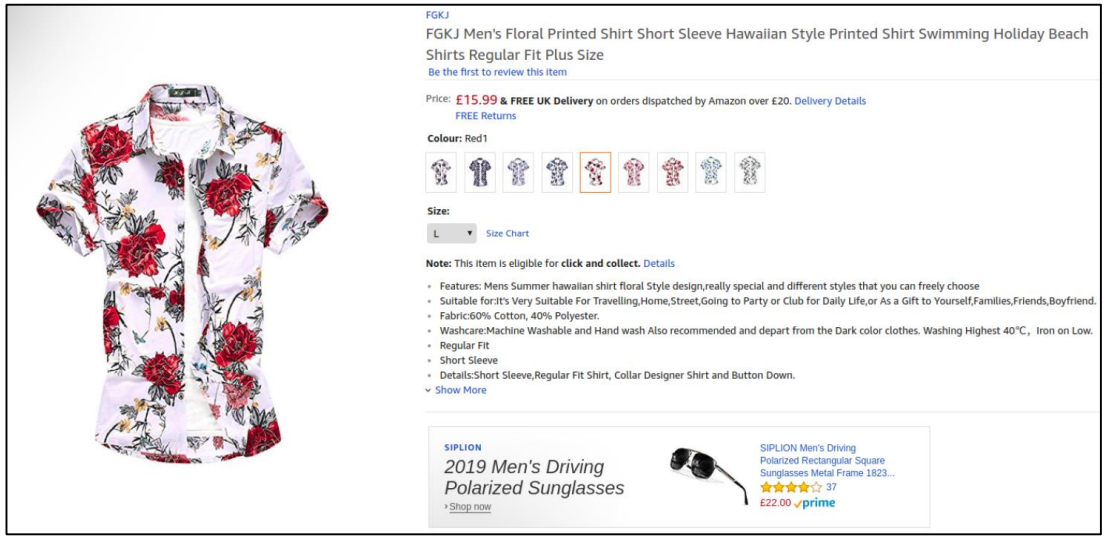
Evaluation- Survey

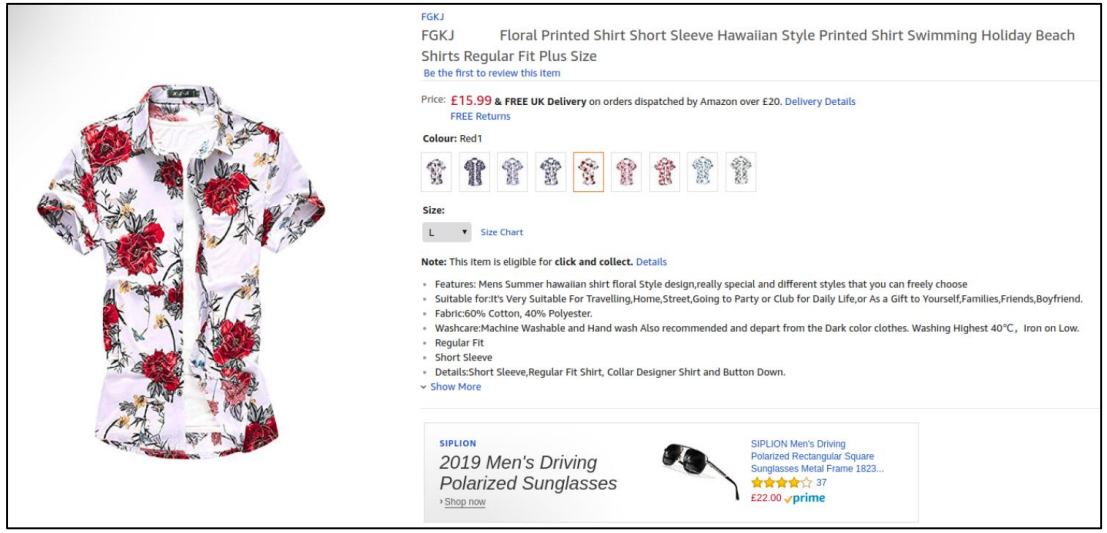


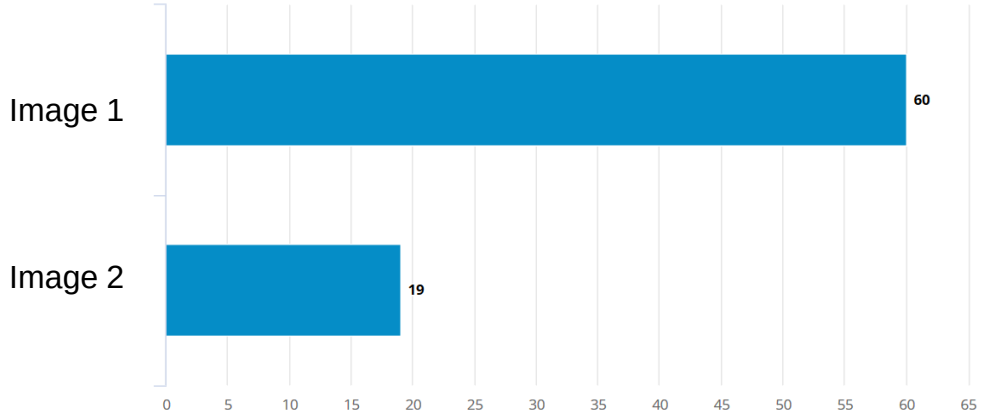
© sebis

Evaluation- Survey

Which detailed page do you prefer to buy this shirt? *



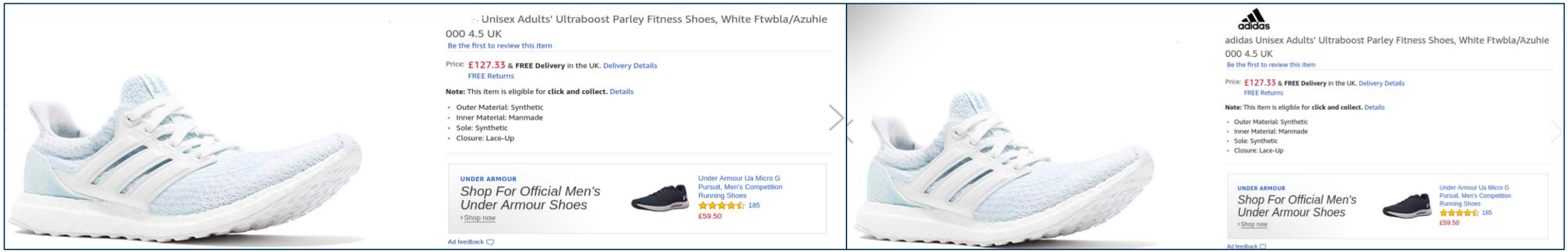


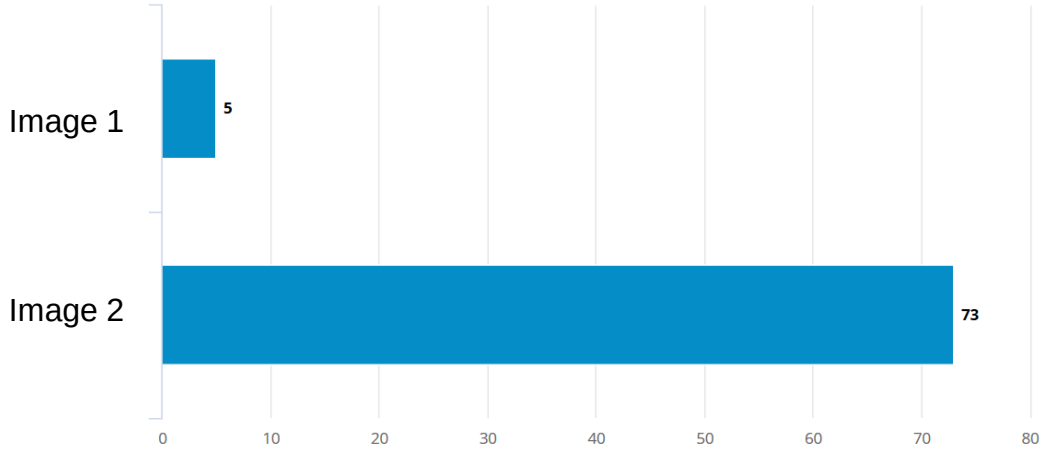


© sebis

Evaluation- Survey

Which of the 2 detailed pages would you rely on to make a buying decision?

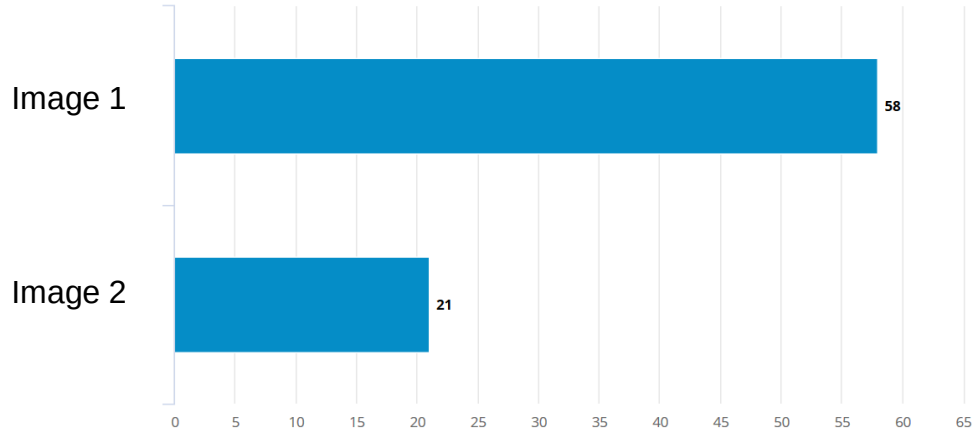




© sebis

Evaluation- Survey

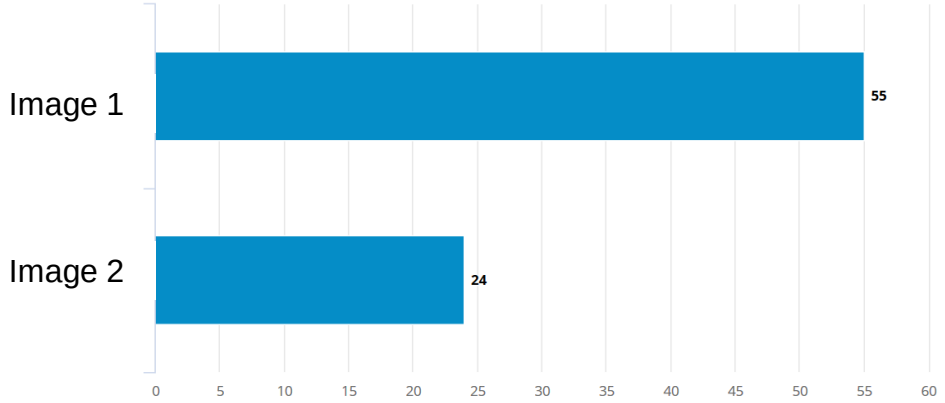
Which detailed page would you rely to make your purchase? *



© sebis

Evaluation- Survey

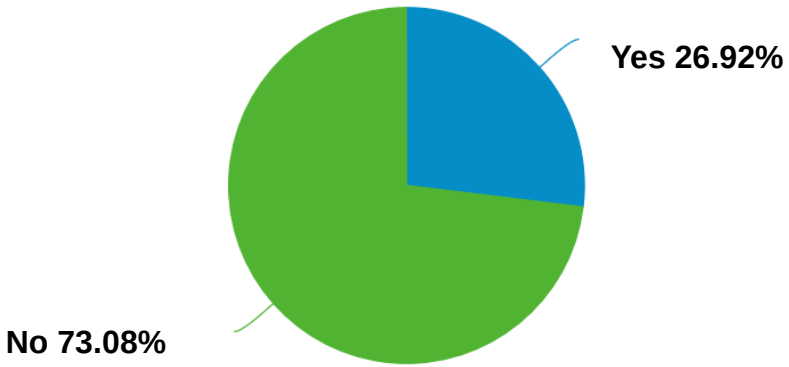
Which detailed page is more helpful for you to make your purchase? *



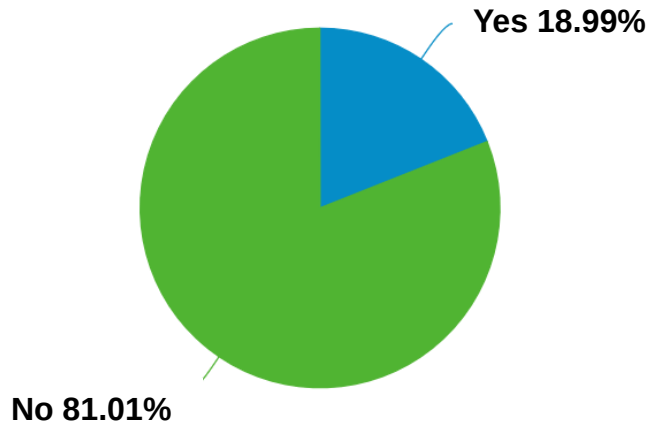
© sebis

Evaluation- Survey

Would you buy a product if the *colour* of the product is missing from the text description?



Would you buy a product if the *brand* of the product is missing from the text description?

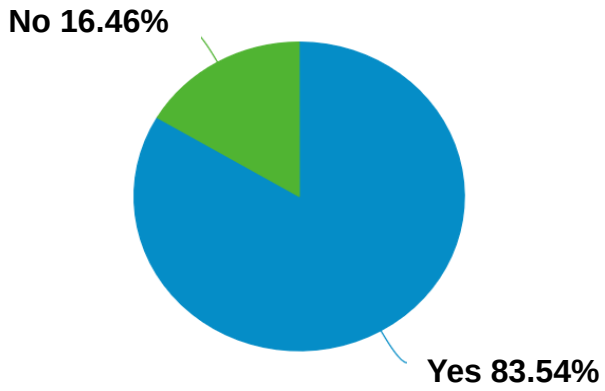


© sebis

Evaluation- Survey

Would you buy a product if the *category* of the product is missing from the text description?

When buying a product would you prefer having the target *gender* the product is for mentioned in the description?



© sebis

Outline

1. Motivation
2. Research Questions
3. Approach
4. Implementation
5. Evaluation
6. Conclusion

© sebis

Research Questions

1	Could Multi-task learning and Transfer Learning perform better than Single task learning on the Amazon Product Catalog dataset?
2	What architecture choices and hyperparameters shall we use in both Muli-task and Transfer learning to obtain good performance?
3	Can Transfer Learning and Multi-task learning be useful in the e-commerce domain to enhance user-experience?

© sebis

Questions?

© sebis

Appendix

© sebis

Appendix

- **Sample**:

Dataset	Input Sample	Label
UK Category	'Holstyle 0.6cm Heel Lift Half Insoles for Loafer....'	Insoles-Comfort
UK Color	'Mini dress with deep V-neck - Pink - 14-16....'	Pink
UK Brand	'Nike Slam Women's Dri-Fit Tennis Skirt - Black....'	Nike
UK Gender	'Cinda Baby Girls Christening Party Dress with Shoes....'	Baby girl
DE Category	'Rockabella Ivy Kleid schwarz/Weiss....'	Kleid
DE Color	'Damen Riemchen Abend Sandaletten High Heels....'	Schwarz
DE Brand	'Wrangler Herren Jeansjacke Auth Western:....'	Wrangler
DE Gender	'Sakkas Azalea Stein gewaschen gestickte Kunstseide Korsett....'	Damen
FR Category	'7 For All Mankind Bootcut, Jeans Femme....'	Jeans
FR Color	'Bloch Criss Cross, Chaussures de Danse....'	Noir
FR Brand	'Adidas FEF H JSY T-Shirt Homme Rouge....'	Adidas
FR Gender	'Chic Feet , Sandales pour femme 37.5....'	Femme

Color: Weiß

Gender: Damen

Category: Jeans

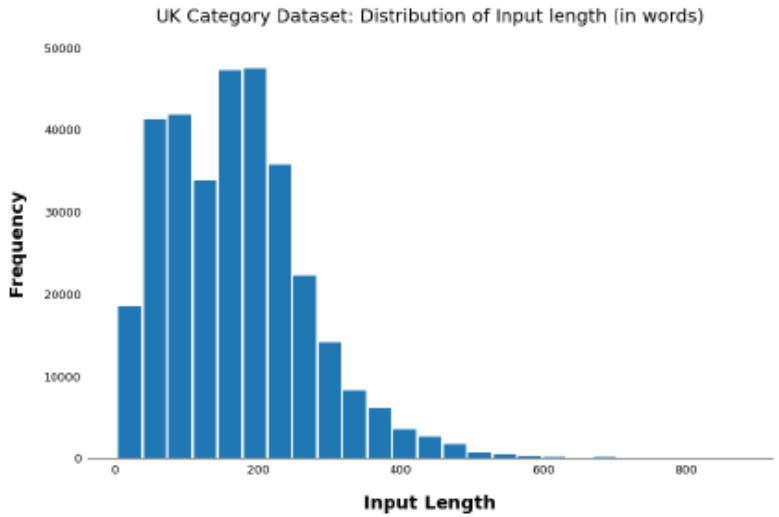
Brand: Saxoline

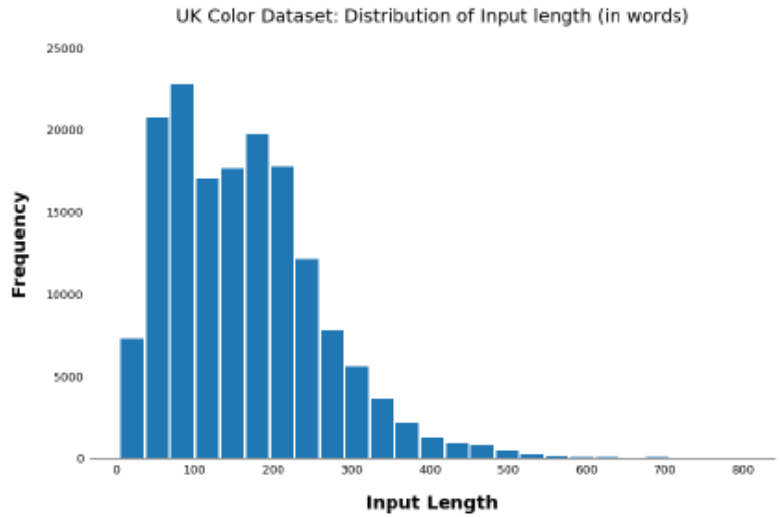
Implementation-Dataset Statistics

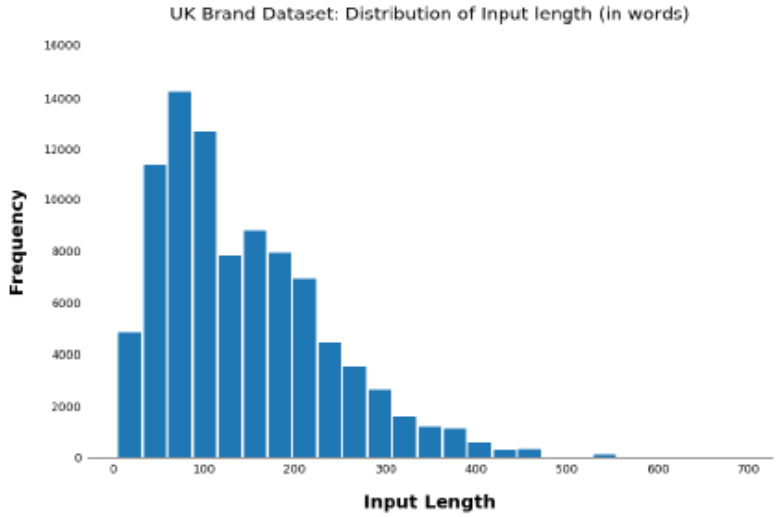
Dataset	Number of Records	Number of unique labels
UK Category	377,121	174
UK Color	190,538	154
UK Brand	303,967	599
UK Gender	387,594	8
DE Category	138,696	103
DE Color	43,343	64
DE Brand	125,362	269
DE Gender	148,090	8
FR Category	284,048	174
FR Color	111,681	120
FR Brand	211,101	498
FR Gender	293,659	6

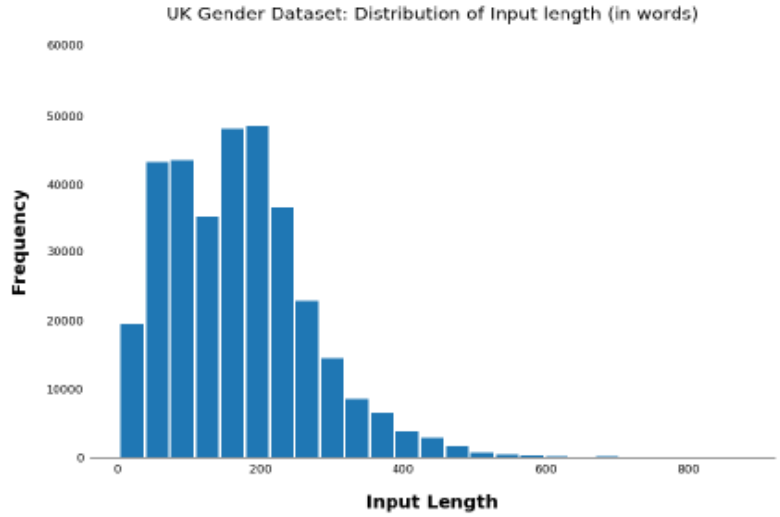
© sebis

Implementation-Dataset Statistics



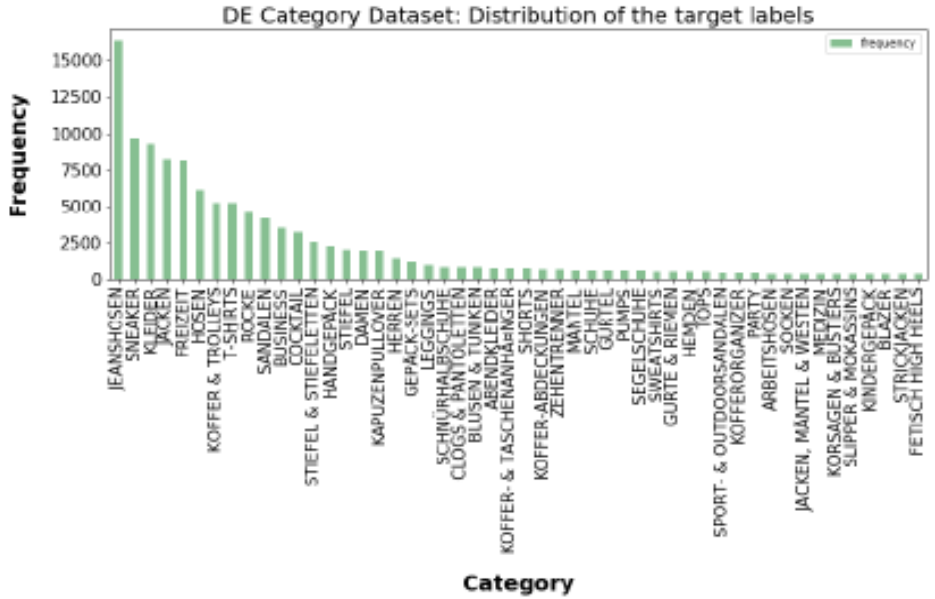


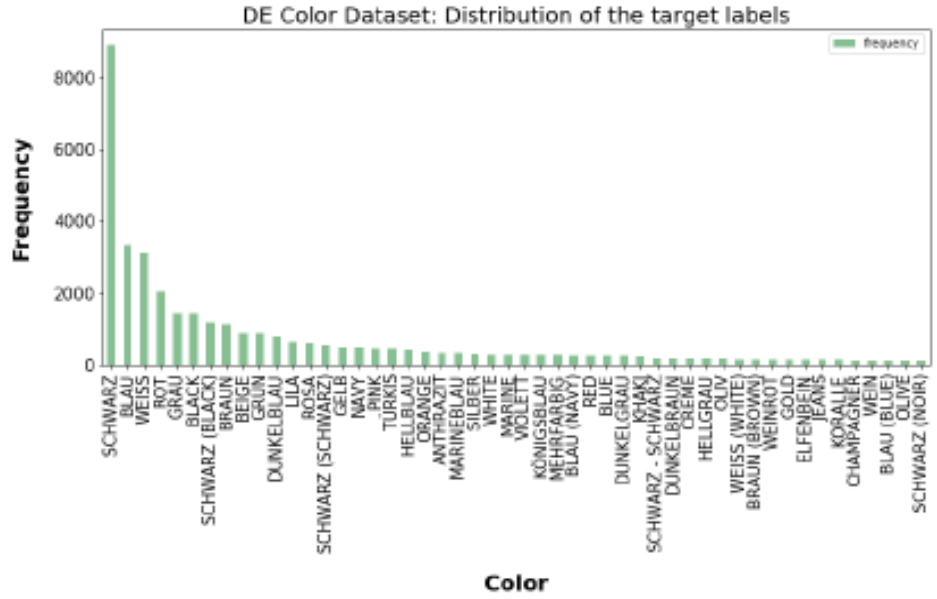


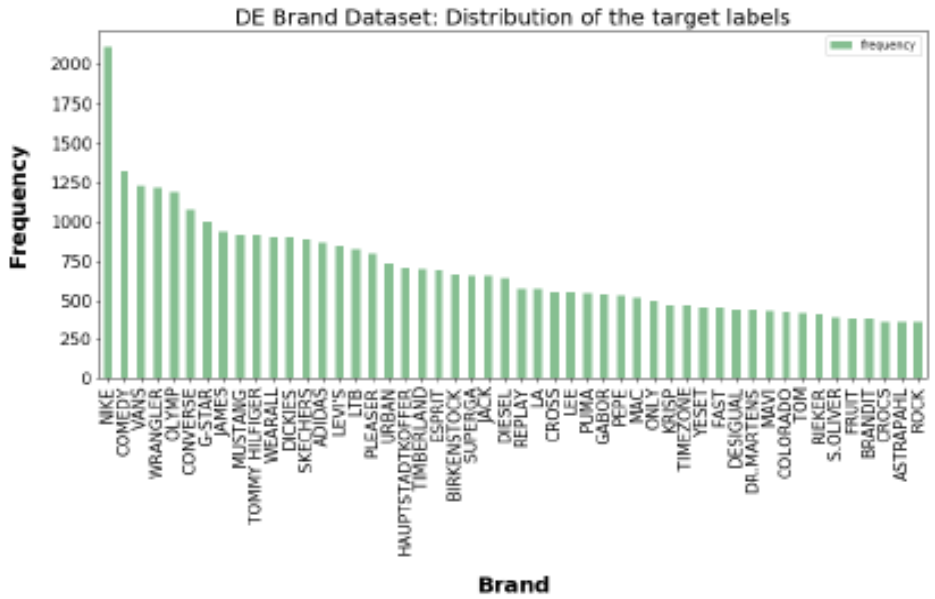


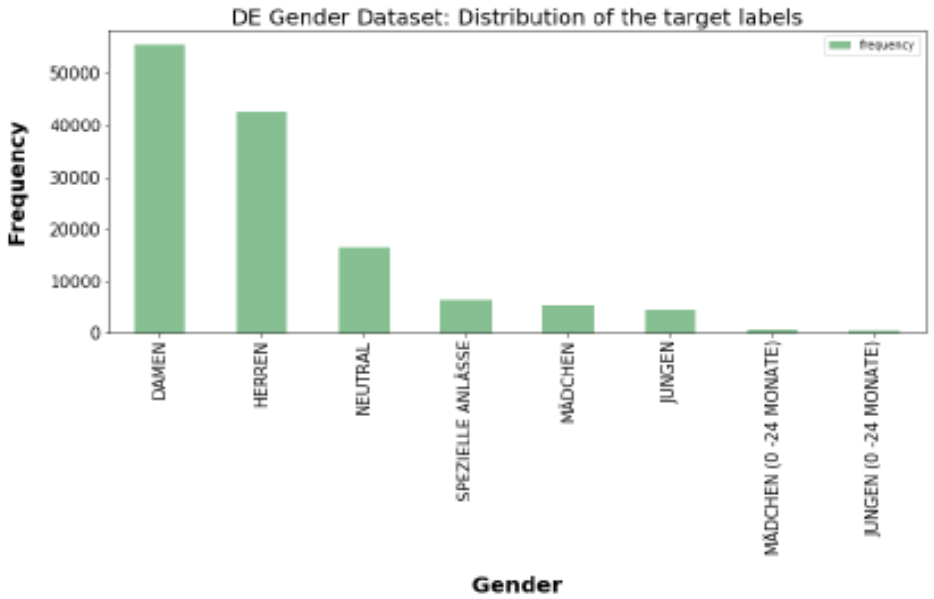
© sebis

Implementation-Dataset Statistics









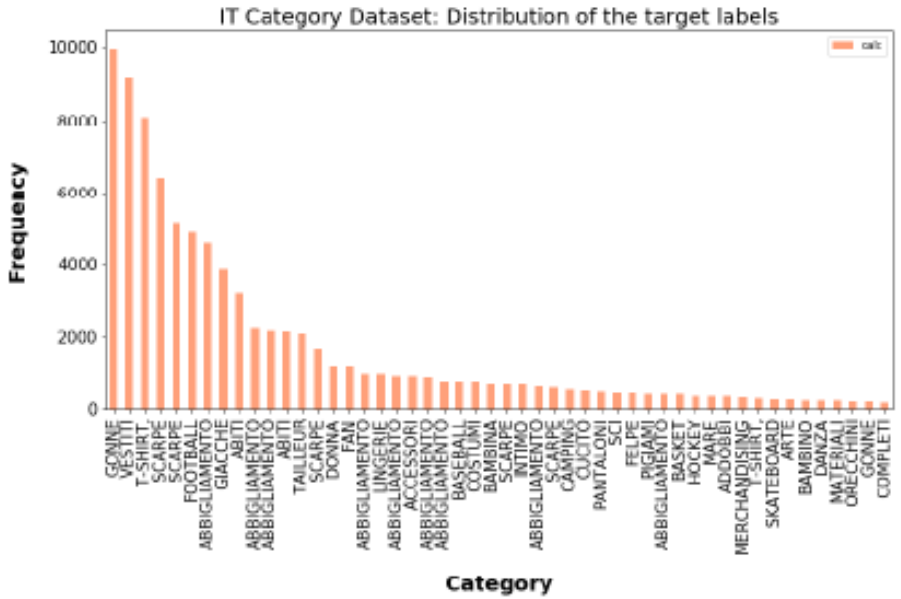
© sebis

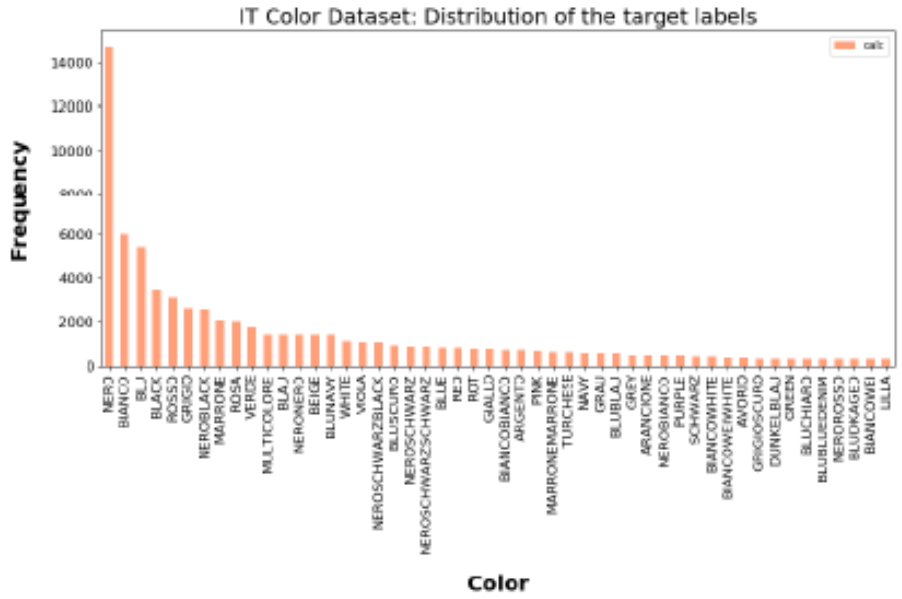
Implementation-Dataset Statistics

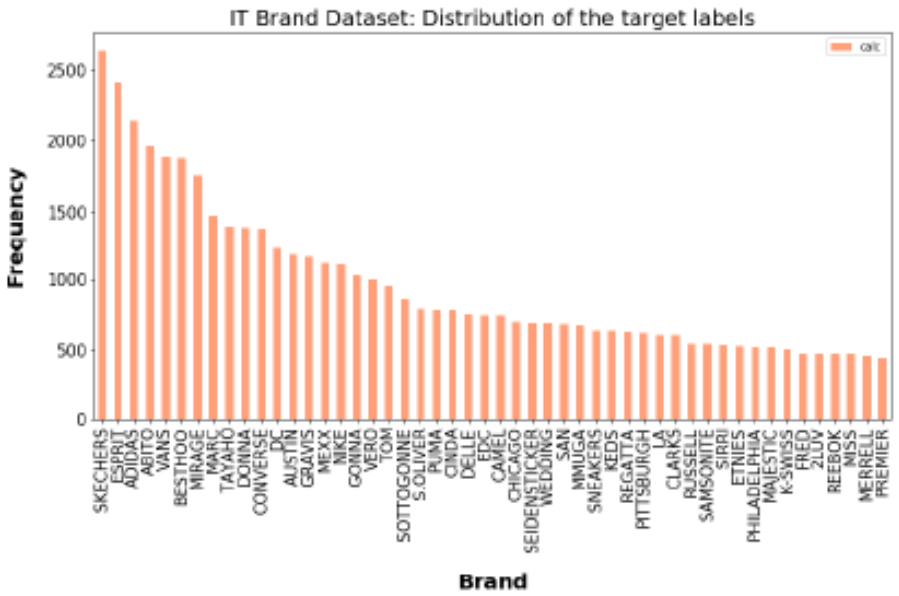
Dataset	Number of Records	Number of unique labels
UK Category	82381	223
UK Color	96203	627
UK Brand	99728	172
UK Gender	113463	8
DE Category	98320	550
DE Color	74846	121
DE Brand	99564	197
DE Gender	102743	7
IT Category	89286	107
IT Color	101218	458
IT Brand	80458	145
IT Gender	87637	5

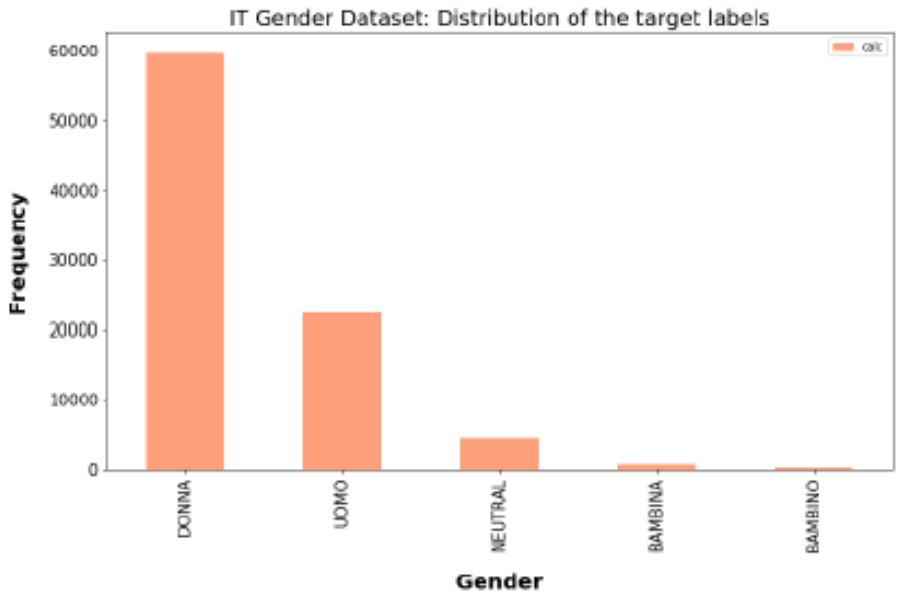
© sebis

Implementation-Dataset Statistics



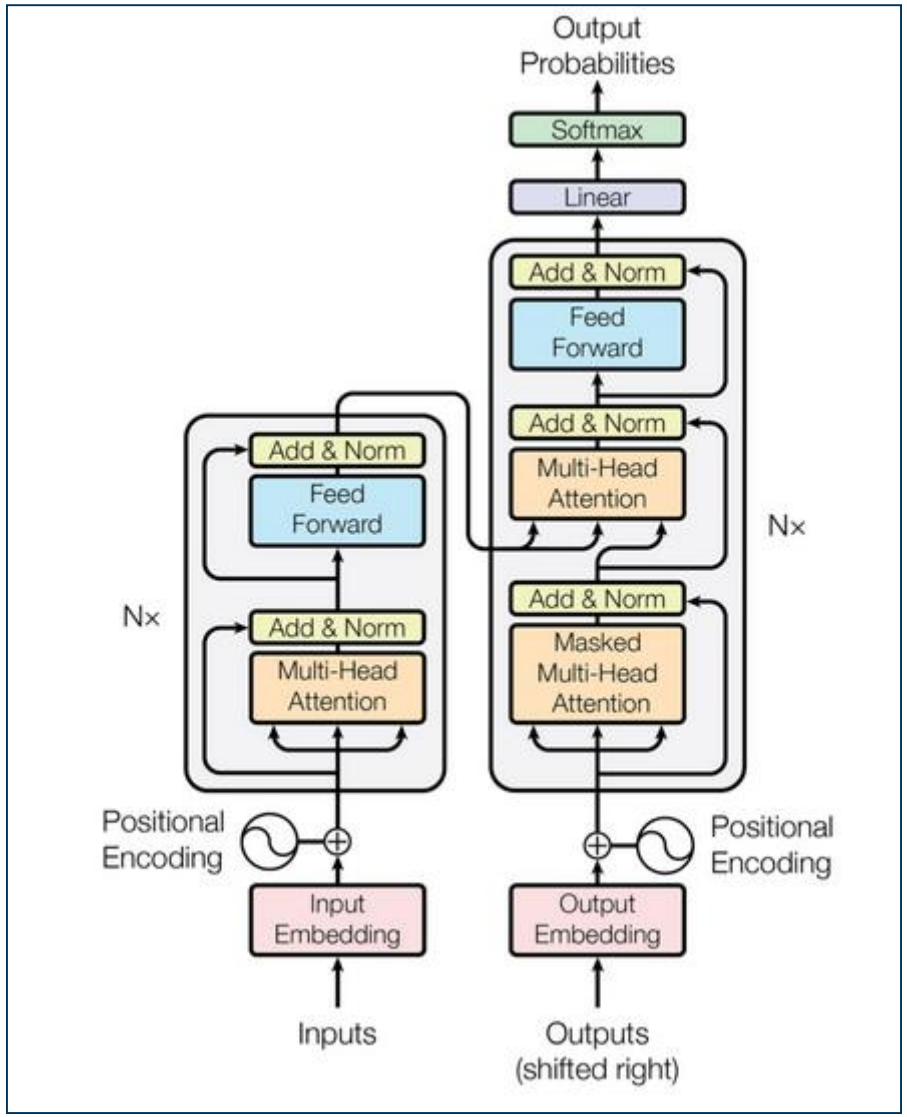






© sebis

Transformer Model



© sebis

Evaluation- Discussion

Modality	Dataset	Single Task		Transfer Learning		Multi-task learning	
		Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
Text	Uk Category	74	56	98	98	98	-
	UK Color	29	13	51	54	56	-
	UK Brand	80	65	100	100	100	-
	UK Gender	93	88	100	99	97	-
	DE Category	79	59	98	96	96	-
	DE Color	51	31	69	58	64	-
	DE Brand	92	73	100	100	100	-
	DE Gender	92	83	100	99	97	-
	FR Category	78	56	98	97	96	-
	FR Color	36	18	60	48	55	-
	FR Brand	93	81	100	100	100	-
	FR Gender	90	81	100	100	96	-

© sebis

Evaluation- Discussion

Modality	Dataset	Single Task		TL - Strategy I		TL - Strategy II	
		Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
Image	Uk Category	88	88	74	74	88	86
	UK Color	60	67	71	71	83	78
	UK Brand	48	50	31	34	55	52
	UK Gender	94	91	90	88	95	92
	DE Category	81	83	45	45	73	84
	DE Color	86	88	47	47	89	88
	DE Brand	57	55	38	38	60	58
	DE Gender	96	96	93	90	96	95
	IT Category	86	84	48	48	83	80
	IT Color	60	67	20	20	58	65
	IT Brand	56	53	65	58	55	66
	IT Gender	96	97	96	96	96	98

© sebis

References

[1] https://www.emarketer.com/Article/Worldwide-Retail-Ecommerce-Sales-Will-Reach-1915-Trillion-This-Year/1014369
[2] https://www.nngroup.com/reports/ecommerce-user-experience/
[3] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu. A Survey on Deep Transfer Learning. arXiv:1808.01974v1, 2018.
[4] J.Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR), 2009.
[5] Y. Zhang and Q. Yang. A Survey on Multi-Task Learning. arXiv:1707.08114v2, 2017
[6] Sebastian Ruder. An overview of Multi-Task Learning in Deep Neural Networks, 2017
[7] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv:1602.07261v2, 2016
[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I Polosukhin. Attention Is All You Need. arXiv:1706.03762v5, 2017
[9]Transformer: A Novel Neural Network Architecture for Language Understanding. 2017. url: https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html
[10] J. Alammar. The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning). 2018. url: http://jalammar.github.io/illustrated-bert/
[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805v2, 2018
[12] Ł. Kaiser, A. N. Gomez, N. Shazeer, A. Vaswani, N. Parmar, L. Jones, and J. Uszkoreit. One Model To Learn Them All. arXiv preprint arXiv:1706.05137v1, 2017.
[13] G. B. Team. tensor2tensor. url: https://github.com/tensorflow/tensor2tensor/blob/master/README.md.

© sebis

Imputation of missing Product Information using Deep Learning

Aamna Najmi

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business Information Systems

Boltzmannstraße 3
85748 Garching bei München

aamna.najmi@mytum.de