

COMP42315 Assignment

1 Question 1

Unlike the logic presented by the code, the overall idea of a web crawler is to first complete crawling a specific page of one paper, then crawl the links to the publications under a topic by for-loop, and then crawl all the links to each topic.

According to the assignment requirements, the title and number of citations must be crawled on the specific page. Considering that the first two topics are the same part of the course and are covered by the same professor, this part is crawled at the same time. And the data is stored in DataFrame for easy processing.

In the process of crawling "Citation", it should be noted that some publications lack the part of "Citation". So if "Citation" exists, the number will be and converted to int and stored. If there is no "Citation", we will store the "Citation" as None, which is more reasonable, because we can not confirm the Citation is 0 or missing.

And in crawling different topic publication links, we should pay attention to the same article appearing in different topics, so we use 'set' to remove duplicates.

When sorting the publications, it needs to consider that the citations of the 25th publication may have the same citation as the following publications. We add year and impact factor here as an aid to sorting, if the citation is the same, the more recent of the year, the more forward, and the greater of the impact factor, the more forward.

A Quadruple Diffusion Convolutional Recurrent Network for Human Motion Prediction	228
Interaction Patches for Multi-Character Animation	201
Real-Time Posture Reconstruction for Microsoft Kinect	150
Simulating Multiple Character Interactions with Collaborative and Adversarial Goals	112
Simulating Competitive Interactions using Singly Captured Motions	112
Simulating Interactions of Avatars in High Dimensional State Space	102
Kinect Posture Reconstruction based on a Local Mixture of Gaussian Process Models	98
Interaction-based Human Activity Comparison	88
Environment-aware Real-Time Crowd Control	79
Real-time Physical Modelling of Character Movements with Microsoft Kinect	68
Posture Reconstruction Using Kinect with a Probabilistic Model	62
Interactive Formation Control in Complex Environments	61
Bi-projection based Foreground-aware Omnidirectional Depth Prediction	60
A Two-Stream Recurrent Network for Skeleton-Based Human Interaction Recognition	60
Interpreting Deep Learning based Cerebral Palsy Prediction with Channel Attention	55
Filtered Pose Graph for Efficient Kinect Pose Reconstruction	55
Emulating Human Perception of Motion Similarity	51
Differential Evolution Algorithm as a Tool for Optimal Feature Subset Selection in Motor Imagery EEG	50
Motion Adaptation for Humanoid Robots in Constrained Environments	50
Spatio-temporal Manifold Learning for Human Motions via Long-horizon Modeling	48
Validation of an Ergonomic Assessment Method using Kinect Data in Real Workplace Conditions	45
Arbitrary View Action Recognition via Transfer Dictionary Learning on Synthetic Training Data	45
DurLAR: A High-fidelity 128-channel LiDAR Dataset with Panoramic Ambient and Reflectivity Imagery for Multi-modal Autonomous Driving Applications	44
SkillVis: A Visualization Tool for Boxing Skill Assessment	44
Multi-layer Lattice Model for Real-Time Dynamic Character Deformation	44

Table 1: Top25 based on Citation

2 Question 2

The logic is the same as that of question 1. When crawling the number of LDO, and analyzing the web page structure, we can find that, except for the embedded youtube video links, all other items can be obtained by looking up the name, just store it as a list and calculate the number of list elements, and the embedded youtube video only needs to calculate the embedded iframe. The sum of the number of the two parts is the

number of LDO items. The same problem as citation is encountered when capturing the impact factor, if there is no impact factor we store it as None.

When dealing with the sorting problem, we use the number of LDO items as the primary criterion, but the number of LDOs is consistent for many papers, so we use the number of citation, year and impact factor as the secondary sorting criteria in turn, where the value of None is ranked last.

The maximum number of LDO is 10 and the minimum is 4 in the papers of LDO number Top25, but there are many papers with the number of 4, and we have selected some of them according to the previous rules. To calculate the average number of LDOs per year, we use the year as the grouping criterion to calculate the average number of LDO items, the smallest being the number of 2 in 2006 and the largest being the number of 5 in 2008, 2009 and 2011.

Interaction Patches for Multi-Character Animation	10
Simulating Multiple Character Interactions with Collaborative and Adversarial Goals	7
Simulating Interactions of Avatars in High Dimensional State Space	7
Resolving Hand-Object Occlusion for Mixed Reality with Joint Deep Learning and Model Optimization	7
DSPP: Deep Shape and Pose Priors of Humans	6
Foreground-aware Dense Depth Estimation for 360 Images	6
Makeup Style Transfer on Low-quality Images with Weighted Multi-scale Attention	6
360 Depth Estimation in the Wild - The Depth360 Dataset and the SegFuse Network	6
A Quadruple Diffusion Convolutional Recurrent Network for Human Motion Prediction	5
Real-Time Posture Reconstruction for Microsoft Kinect	5
Interaction-based Human Activity Comparison	5
A Two-Stream Recurrent Network for Skeleton-Based Human Interaction Recognition	5
Interpreting Deep Learning based Cerebral Palsy Prediction with Channel Attention	5
Topology Aware Data-Driven Inverse Kinematics	5
Physically-based Character Control in Low Dimensional Space	5
Angular Momentum Guided Motion Concatenation	5
Fast Accelerometer-Based Motion Recognition with a Dual Buffer Framework	5
Data-Driven Crowd Motion Control with Multi-touch Gestures	5
A Unified Deep Metric Representation for Mesh Saliency Detection and Non-rigid Shape Matching	5
Formation Control for UAVs Using a Flux Guided Approach	5
Semantics-STGCNN: A Semantics-guided Spatial-Temporal Graph Convolutional Network for Multi-class Trajectory Prediction	5
Single Sketch Image based 3D Car Shape Reconstruction with Deep Learning and Lazy Learning	5
Simulating Competitive Interactions using Singly Captured Motions	4
Kinect Posture Reconstruction based on a Local Mixture of Gaussian Process Models	4
Real-time Physical Modelling of Character Movements with Microsoft Kinect	4

Table 2: Top25 based on LDO items

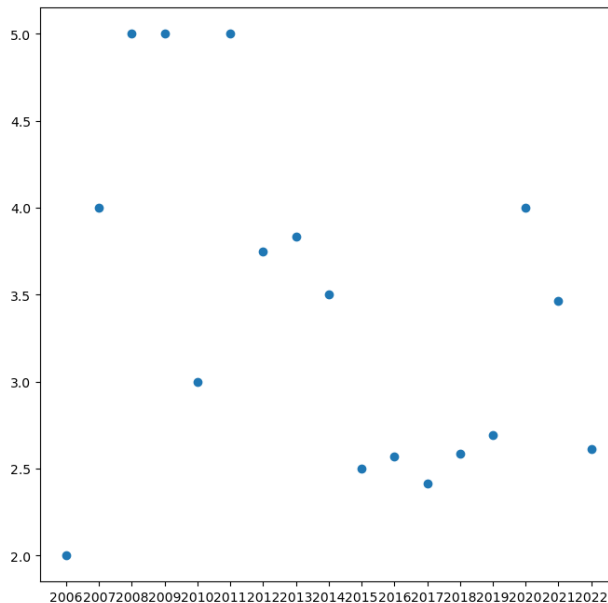


Figure 1: Average number of LDO items against the year

3 Question 3

The top ten 'highly cited' authors are obtained by grouping the authors and calculating sum of the number of citations. The same publication appears under a different topic, so it is essential to match the relationship between the topic and the publication. Because of the co-authors involved, the article with one author will be deleted. We use the first author as the primary author and the others as the co-author, so we need to delete the first author, and the rest are grouped according to topics as the co-author.

The data of articles with one author need to be removed, then calculate the number of authors appearing under each topic, and get the most popular topic by sorting the number of authors.

3a. Hubert's publication has the highest citations at over 3,000; three have more than 1,000, and the lowest citations among the top ten authors is 220. The top ten authors have the most citations in Character Animation, with over 1000, followed by 3D Reconstruction and Motion Analysis. The top ten authors with the most collaborations are Character Animation, Interaction Modelling and Motion Analysis, all with more than 40 citations.

3b. feature information includes venue, year, publisher, topic, impact factor, number of authors and number of LDOs. According to the impact factor segmentation provided, we convert the impact factor and the number of LDOs into categorical data. The features were analyzed by ANOVA. Then, year, publisher and number of LDOs are the top three features that have the highest impact on the citation, and the bar charts can also be a supplementary illustration.

3c. The data results show that only the papers published in Journal have impact factors, the total number of impact factors is over 200, and the average impact factor of each paper is also over 3.5.

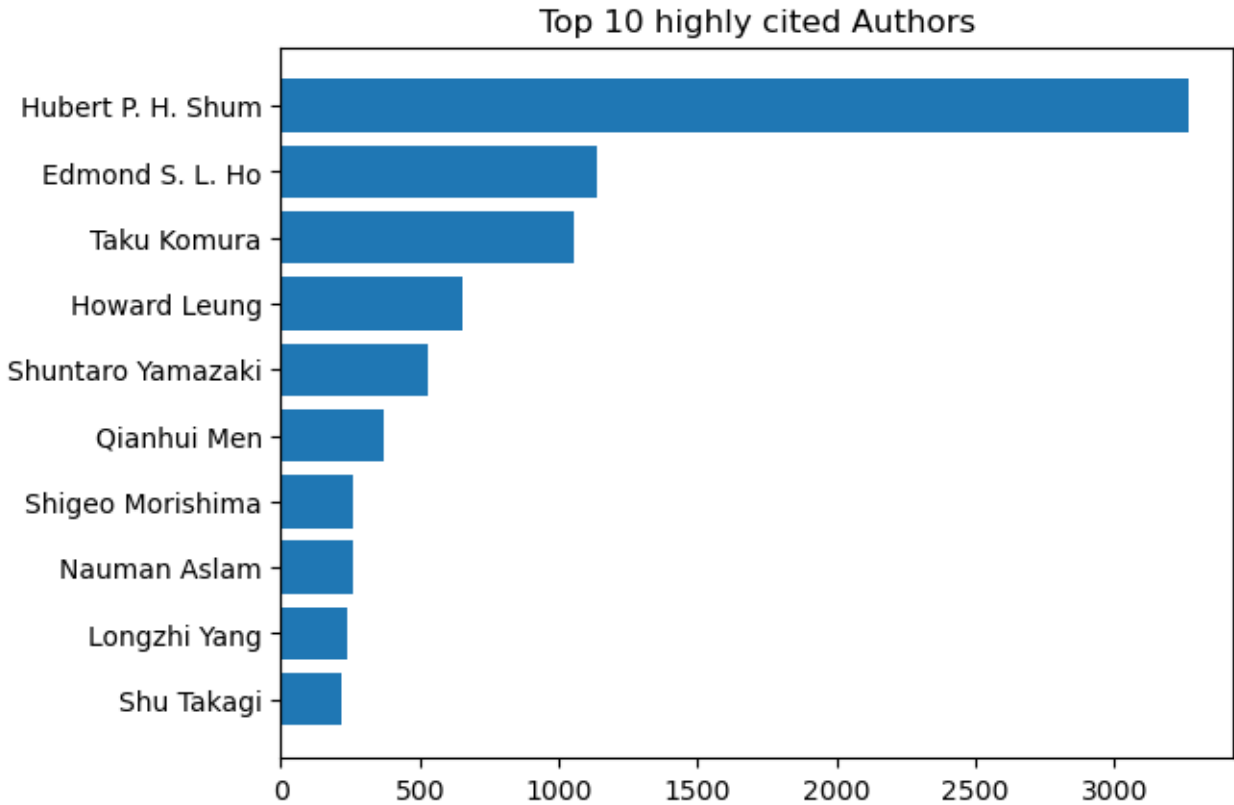


Figure 2: Top 10 highly cited Authors

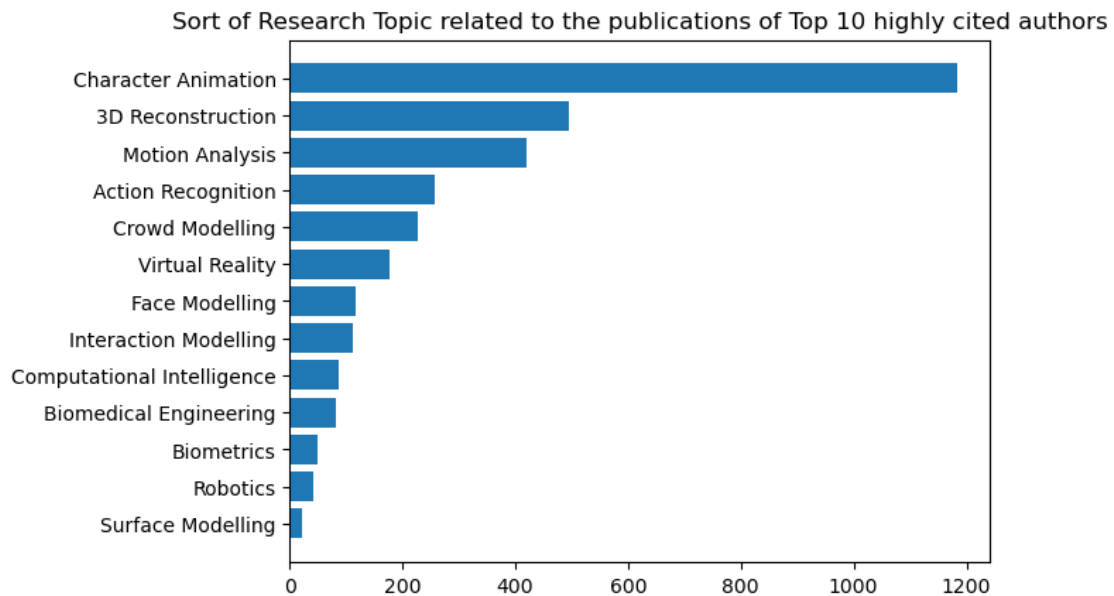


Figure 3: Sort of Research Topics

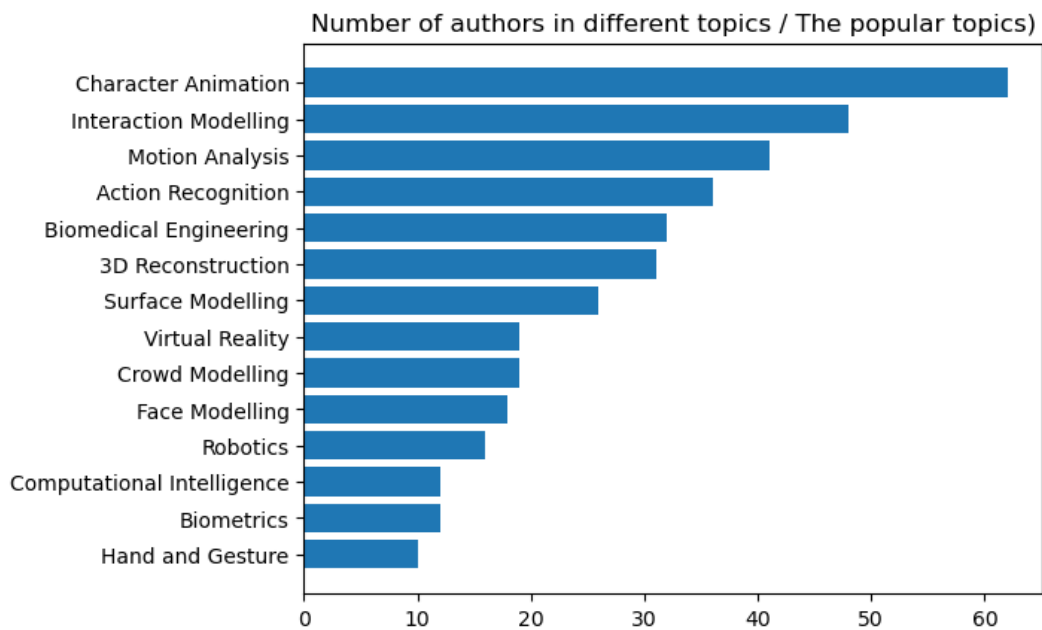


Figure 4: The Most Popular Topics

4 Question 4

The DAG in the tree-augmented Naive Bayesian network shows that covid level affects all symptoms, followed by headache which affects the other three symptoms: loss of smell, chest pain and diarrhea, chest pain affects two symptoms: hoarseness and confusion, loss of appetite affects abdominal pain and fever, and abdominal pain affects cough and fatigue. loss of smell, hoarseness, short of breath, fever and cough do not affect other symptoms. And The remaining symptoms will only affect one other symptom. The Conditional Probability Table quantifies the findings of the DAG. The structure correlation scores show that the Accuracy score reaches 0.9, but all other scores are 0 or negative. Therefore, although the model prediction accuracy is high, the overall evaluation is not so good which means that the current model is not good and needs to continue to be adjusted.

The results of the Complex Bayesian Network show that covid level affects all symptoms, but the symptoms do not affect each other. The structure correlation scores are consistent with the results of the tree-augmented Naive Bayesian network, so although the model has high prediction accuracy, and comprehensive evaluation performs not well, the current model needs to continue to be adjusted.

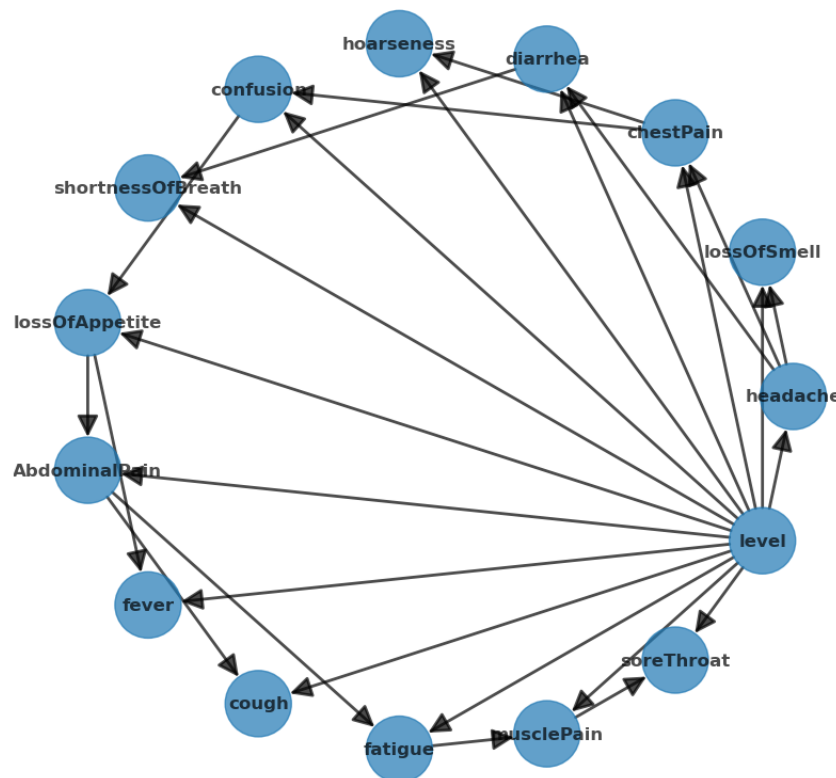


Figure 5: DAG of Tree-Augmented Naive Bayes Network

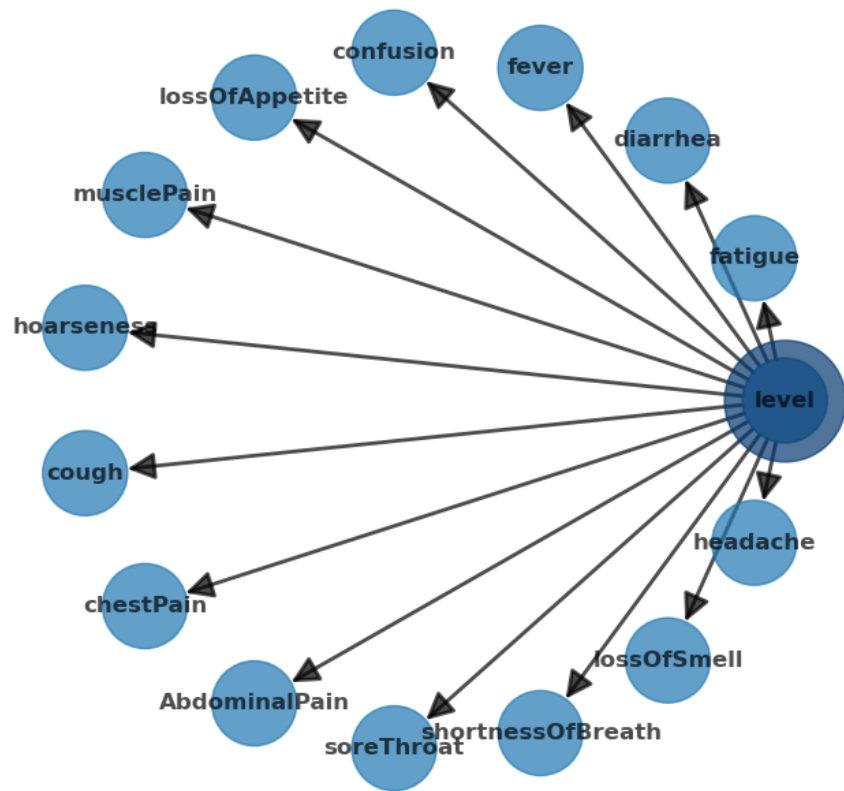


Figure 6: DAG of Complex Bayesian Network