An Open Source Repository of Explainable Artificial Intelligence Projects

Sohee Cho

Explainable Artificial Intelligence Center
KAIST

Explainable AI Program in Korea

Goal Human-level Learning and Inference to overcome the limitations of Deep Neural Networks

AS-IS (Deep Learning)

- It is hard to know the decision, so called **Blackbox** model
- It does not work well when we **do not have enough training data**

TO-BE (Human-level Learning/Inference)

- **Explainable learners** which can provide the reasons of decisions
- Learning explainable models even with **data deficient environment**

Fund Institute of Information & Communication Technology Promotion (IITP) under Ministry of Science and ICT (MSICT) as part of Innovative Growth Engine Project

Period July 2017 ~ December 2021 (54 months)

Project Organization

Partners

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
Technische Universität Berlin
UC DAVIS UNIVERSITY OF CALIFORNIA
JOHNS HOPKINS UNIVERSITY
IBM WATSON
MICHIGAN STATE UNIVERSITY
Penn
MAX PLANCK GESELLSCHAFT
Microsoft Research
MIT
UCLA
SRI International
Carnegie Mellon University

AI algorithms and applications that provide explanations

http://www.openXAI.org/

SW release and Technology promotion

Explainable AI Research Center

AITRICS
KAIST 1971
UNIST ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY 2009
KOREA UNIVERSITY 1905
YONSEI UNIVERSITY 1885
VERITAS LUX MEA

Technology & DB Cooperation

Technology Sharing
Competition

kaggle

International Finance Trading / Medical Big Data Competition

Obtain medical data & Evaluate usability

Medical application

YONSEI UNIVERSITY 세브란스 SEVERANCE
UNIST 빅데이터연구센터

Obtain financial data & Evaluate usability

Financial application

Network Synergy THINKPOOL
UNIST 국제에너지트레이딩연구센터

Demand for technology
Technology transfer & commercialization

Enterprises with intention of technology transfer

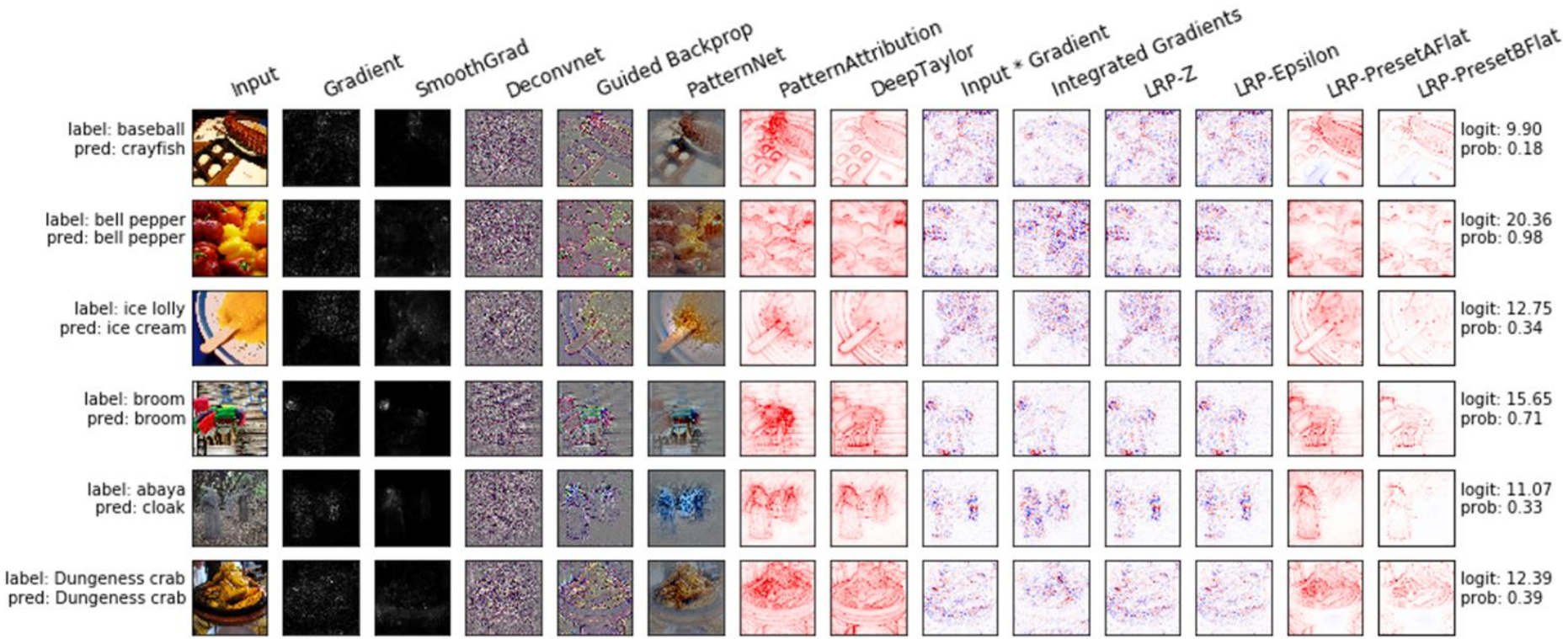
NAVER
AI ZEN AI for Better Finance
posco
BNK 경남은행


xaic
HOME
ABOUT
NEWS&INFO
RESEARCH
OPEN SOURCE
EVENT
CONTACT
Wokrshop
GITHUB
RELATED PROJECT
COMMUNITY
ICCV 2019
Seoul, Korea
2019 ICCV Workshop on
Interpretating and Explaining Visual Artificial Intelligence Models

https://Openxai.org

http://xai.unist.ac.kr/opensource/relatedproject/

iNNvestigate neural networks



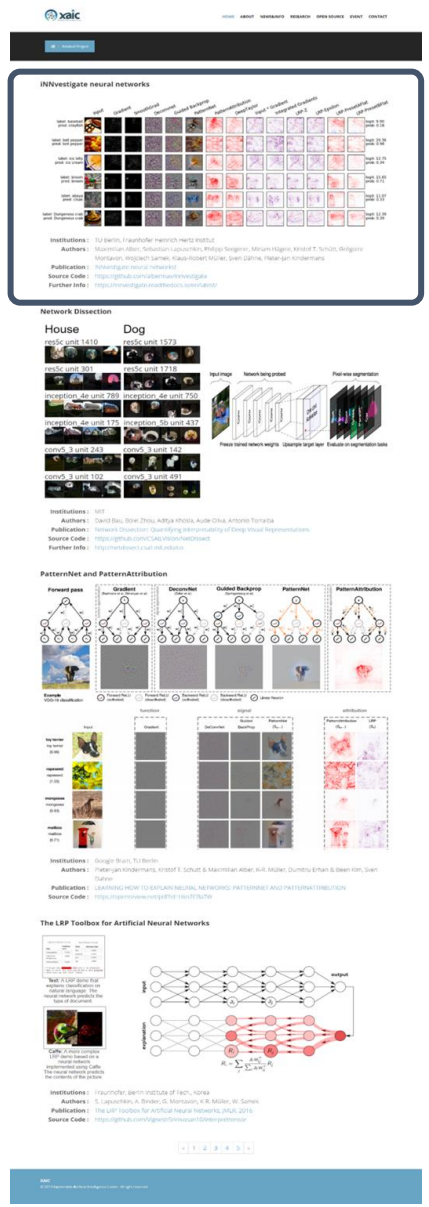
Institutions : TU Berlin, Fraunhofer Heinrich Hertz Institut

Authors : Maximilian Alber, Sebastian Lapuschkin, Philipp Seegerer, Miriam Hägele, Kristof T. Schütt, Grégoire Montavon, Wojciech Samek, Klaus-Robert Müller, Sven Dähne, Pieter-Jan Kindermans

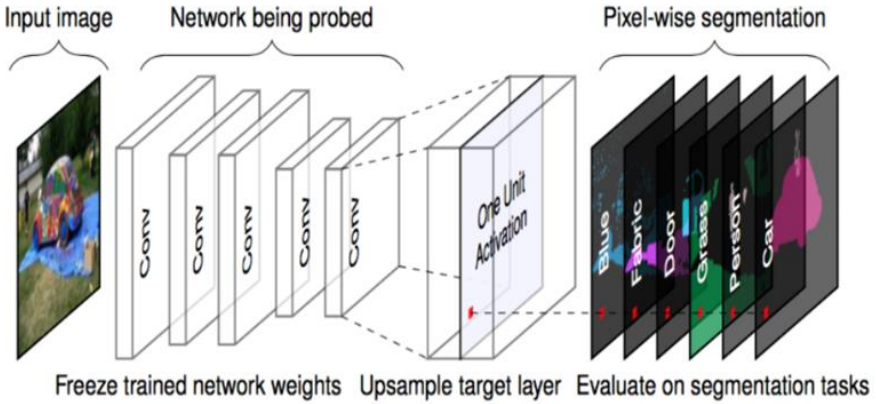
Publication : iNNvestigate neural networks!

Source Code : https://github.com/albermax/innvestigate

Further Info : https://innvestigate.readthedocs.io/en/latest/



Network Dissection



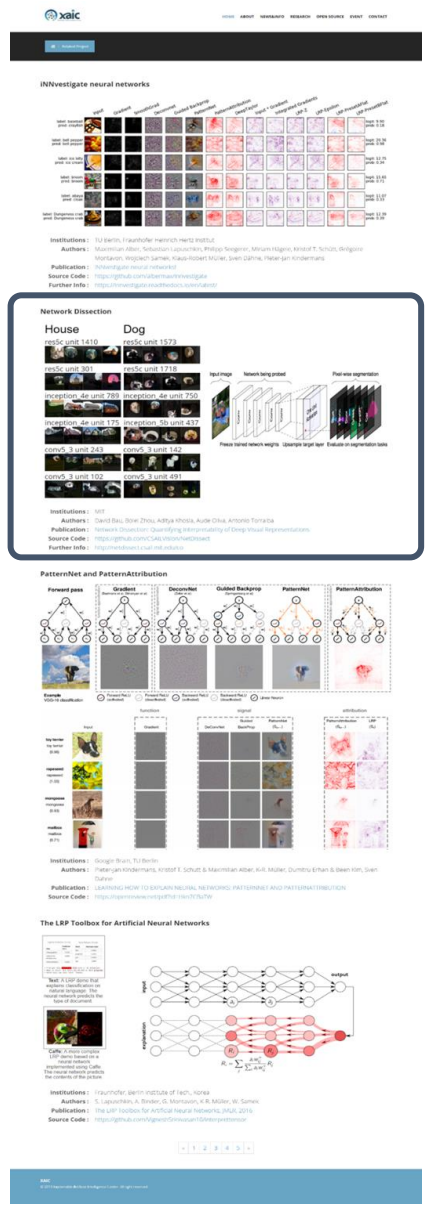
Institutions : MIT

Authors : David Bau, Bolei Zhou, Aditya Khosla, Aude Oliva, Antonio Torralba

Publication : Network Dissection: Quantifying Interpretability of Deep Visual Representations

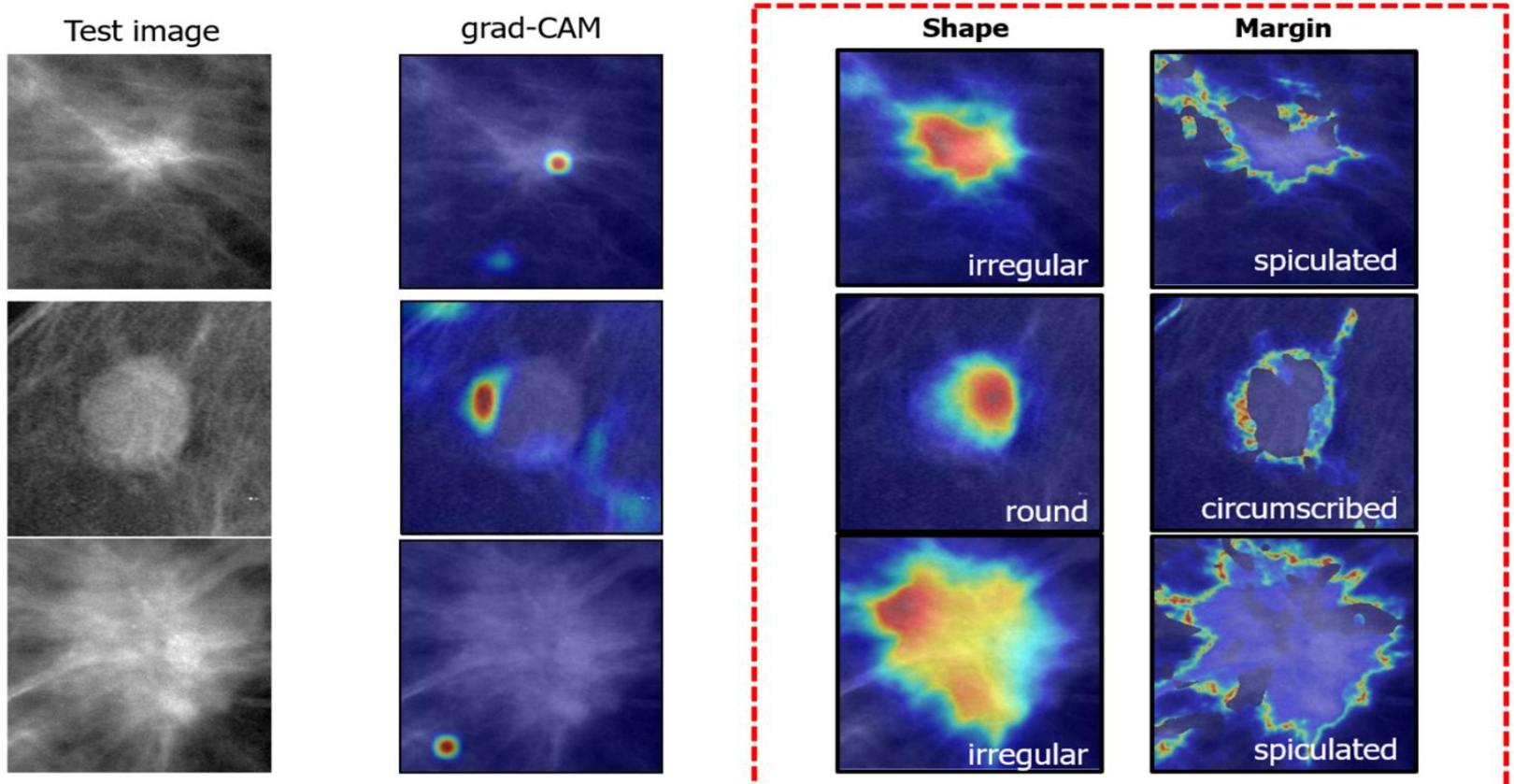
Source Code : https://github.com/CSAILVision/NetDissect

Further Info : http://netdissect.csail.mit.edu/co

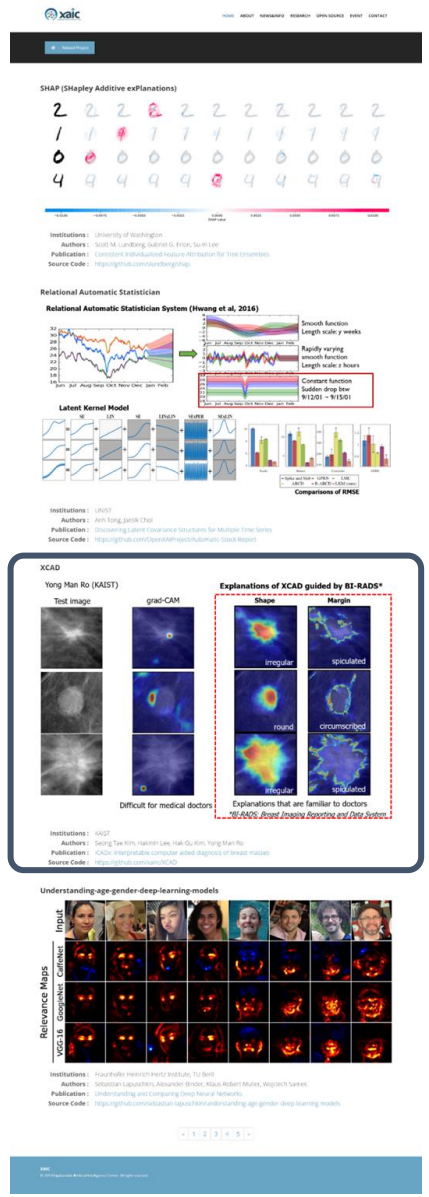


XCAD

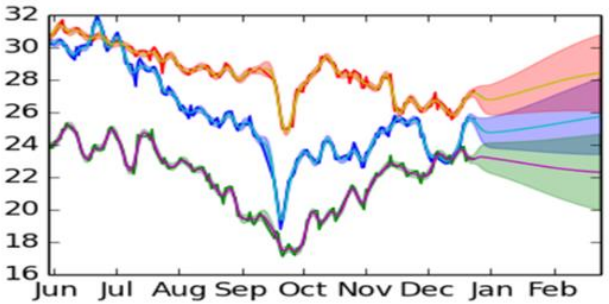
Yong Man Ro (KAIST)

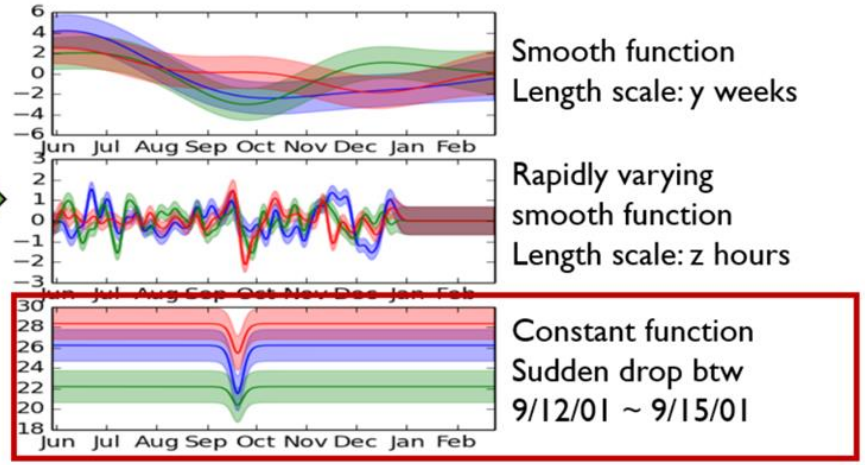


Institutions : KAIST
Authors : Seong Tae Kim, Hakmin Lee, Hak Gu Kim, Yong Man Ro
Publication : ICADx: Interpretable computer aided diagnosis of breast masses
Source Code : https://github.com/xairc/XCAD

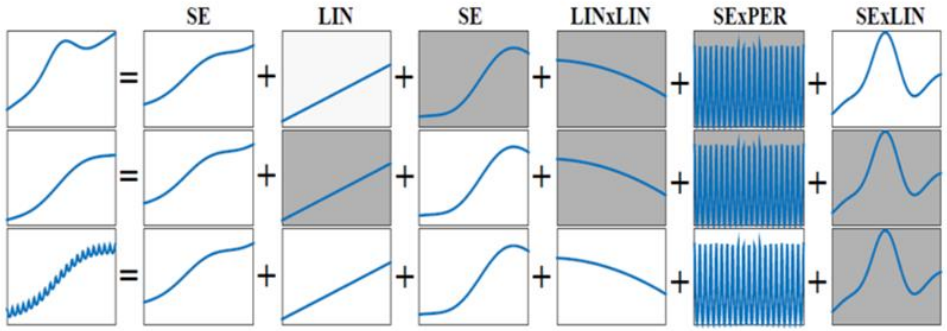


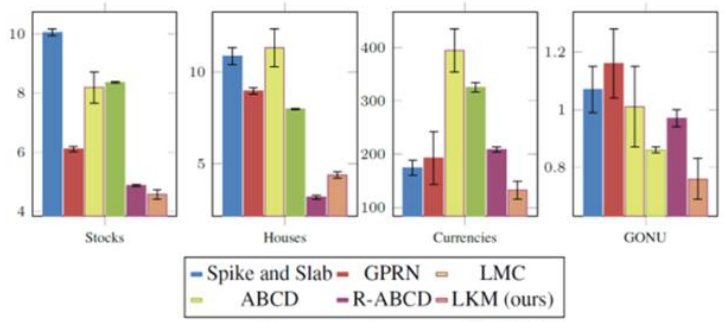
Relational Automatic Statistician





Latent Kernel Model





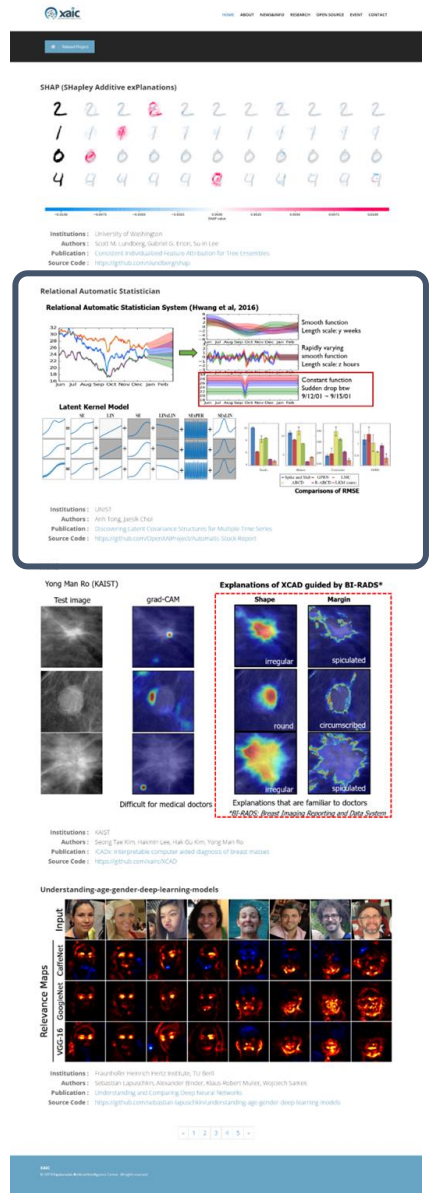
Comparisons of RMSE

Institutions : UNIST

Authors : Anh Tong, Jaesik Choi

Publication : Discovering Latent Covariance Structures for Multiple Time Series

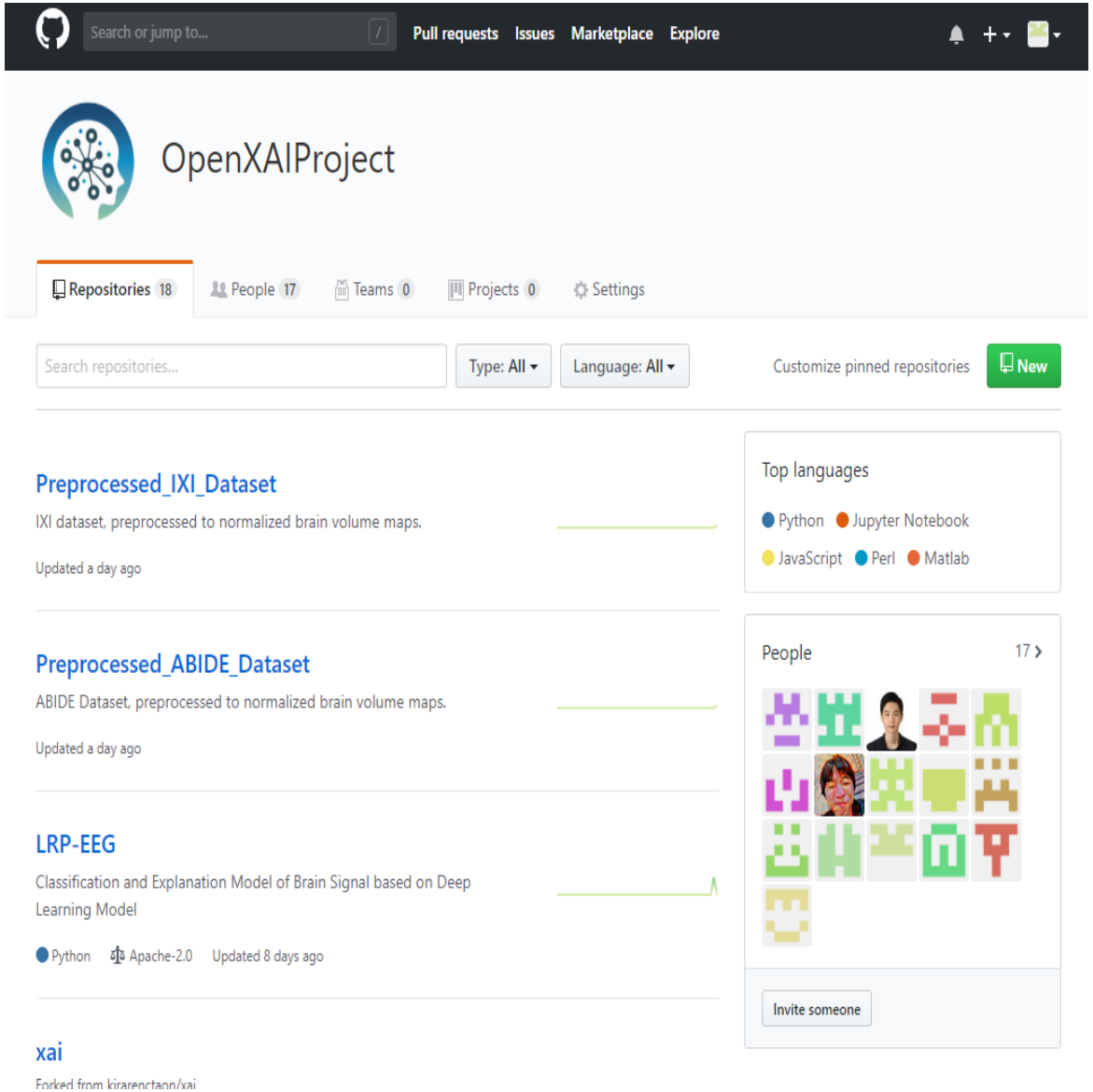
Source Code : https://github.com/OpenXAIProject/Automatic-Stock-Report

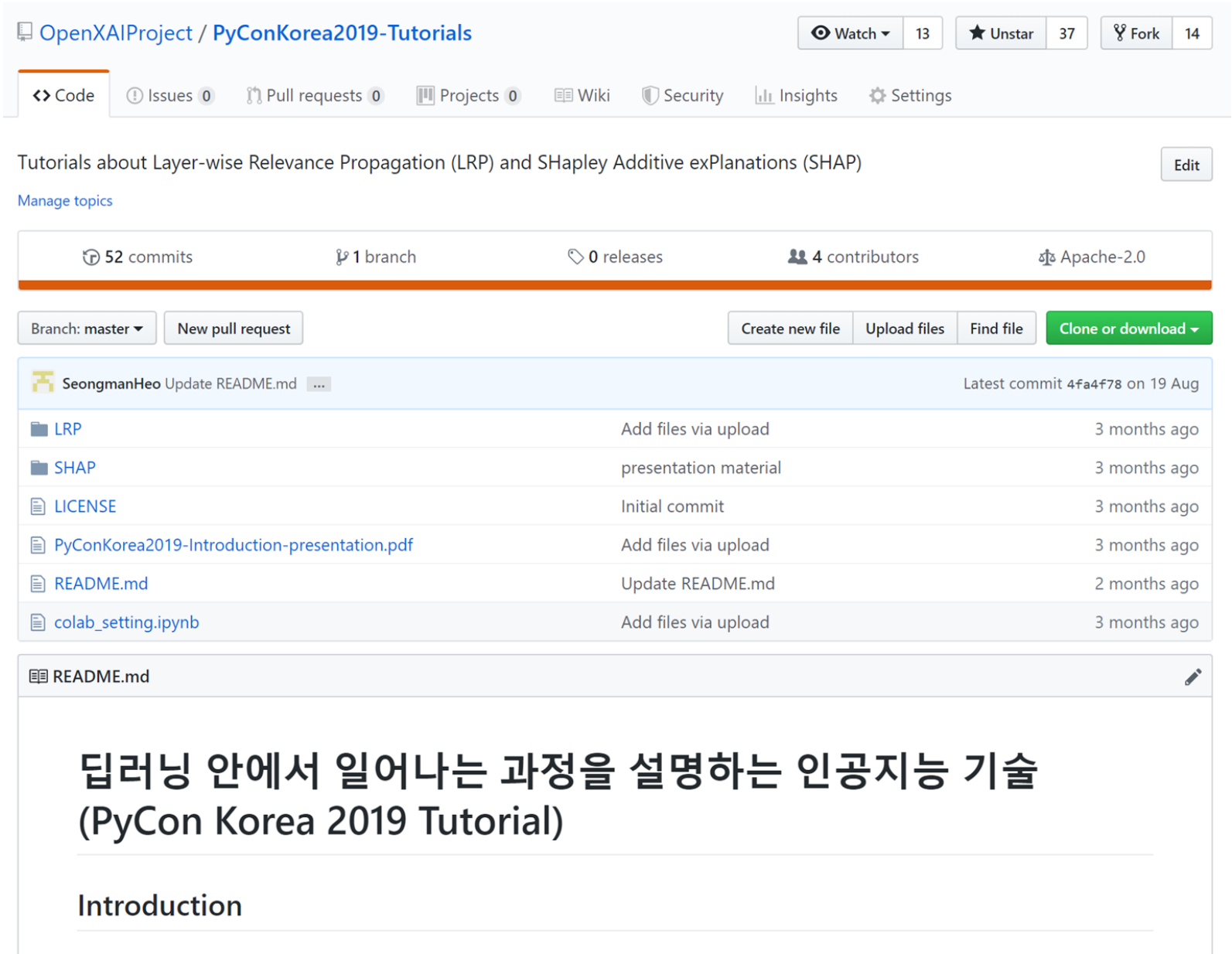


	Project Title	Institutions	Authors	Publication_title	Sourcecode
1	Principles of Explanatory Debugging to Personalize Interactive Machine Learning	Oregon State, City University London	T. Kulesza, M. Burnett, W-K. Wong and S. Stumpf	Principles of Explanatory Debugging to Personalize Interactive Machine Learning, IUI, 2015	https://github.com/fflewddur/IMLPlayground
2	Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model	MIT, U of Washington, Columbia	B. Letham, C. Rudin, T. McCormick and D. Madigan	Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model, Annals of Applied Statistics, 2015	https://github.com/nlarusstone/corels
3	Explaining Recurrent Neural Network Predictions in Sentiment Analysis	Fraunhofer, TU Berlin, Korea University, Max	L. Arras, G. Montavon, K-R. M체ller and W. Samek	Explaining Recurrent Neural Network Predictions in Sentiment Analysis, EMNLP, 2017	https://github.com/ArrasL/LRP_for_LSTM
4	Why Should I Trust You?: Explaining the Predictions of Any Classifier""	U of Washington	M. T. Ribeiro, S. Singh, S. and C. Guestrin	Why Should I Trust You?: Explaining the Predictions of Any Classifier, KDD, 2016	https://github.com/marcotcr/lime
5	Multimedia Event Detection and Recounting	SRI-International Sarnoff, U. of Massachusett	H. Cheng et. al.	SRI-Sarnoff AURORA at TRECVID 2014 - Multimedia Event Detection and Recounting	https://www.nist.gov/itl/iad/mig/tools
6	Examples are not Enough, Learn to Criticize! Criticism for Interpretability	Allen Institute, UT Austin, UIUC	B. Kim, R. Khanna, S. Koyejo	Examples are not Enough, Learn to Criticize! Criticism for Interpretability, NIPS, 2016	https://github.com/BeenKim/MMD-critic
7	Learning AND-OR Templates for Object Recognition and Detection	UC Los Angeles	Z. Si and S. Zhu	Learning AND-OR Templates for Object Recognition and Detection, TPAMI, 2013	http://www.stat.ucla.edu/~zzsi/AOT.html
8	Human-level concept learning through probabilistic program introduction	New York University, U of Toronto, MIT	B. H. Lake, R. Salakhutdinov, and J. B. Tenenbaum	Human-level concept learning through probabilistic program introduction, Science, 2015	https://github.com/brendenlake/BPL
9	Generating Visual Explanations	UC Berkeley, Max Planck Institute for Informa	L. A. Hendricks, Z. Akata, M. Rohrbach, J. Donahue, B. Schiele, and T. Darrell	Generating Visual Explanations, ECCV 2016	https://github.com/LisaAnne/ECCV2016/tree/ECCV2016/examples/ECCV2016
10	The Automatic Statistician	Cambridge, MIT	J. R. Lloyd, D. Duvenaud, R. Grosse, J. B. Tenenbaum and Z. Ghahramani	Automatic Construction and Natural-Language Description of Nonparametric Regression Models, AAAI, 2014	https://github.com/jamesrobertlloyd/gpss-research

	Project Title	Institutions	Authors	Publication_title	Sourcecode
11	Explaining How a Deep Neural Network Trained with End-to-End Learning Steers a Car	NVIDIA, New York University, Google	B. Mariusz et. al.	Explaining How a Deep Neural Network Trained with End-to-End Learning Steers a Car	https://github.com/maxritter/SDC-End-to-end-driving
12	The LRP Toolbox for Artificial Neural Networks	Fraunhofer, Berlin Institute of Tech., Korea	S. Lapuschkin, A. Binder, G. Montavon, K-R. M체ller, W. Samek	The LRP Toolbox for Artificial Neural Networks, JMLR, 2016	https://github.com/VigneshSrinivasan10/interprettensor
13	PatternNet and PatternAttribution	Google Brain, TU Berlin	Pieter-Jan Kindermans, Kristof T. Schutt & Maximilian Alber, K-R. M체ller, Dumitru Erhan & Been Kim, Sven Dahne	LEARNING HOW TO EXPLAIN NEURAL NETWORKS: PATTERNNET AND PATTERNATTRIBUTION	https://openreview.net/pdf?id=Hkn7CBaTW
14	Network Dissection	MIT	David Bau, Bolei Zhou, Aditya Khosla, Aude Oliva, Antonio Torralba	Network Dissection: Quantifying Interpretability of Deep Visual Representations	https://github.com/CSAILVision/NetDissect
15	iNNvestigate neural networks	TU Berlin, Fraunhofer Heinrich Hertz Institut	Maximilian Alber et al	iNNvestigate neural networks!	https://github.com/albermax/innvestigate
16	Understanding-age-gender-deep-learning-models	Fraunhofer Heinrich Hertz Institute, TU Berli	Sebastian Lapuschkin, Alexander Binder, Klaus-Robert Muller, Wojciech Samek	Understanding and Comparing Deep Neural Networks	https://github.com/sebastian-lapuschkin/understanding-age-gender-deep-learning-models
17	XCAD	KAIST	Seong Tae Kim, Hakmin Lee, Hak Gu Kim, Yong Man Ro	ICADx: Interpretable computer aided diagnosis of breast masses	https://github.com/xairc/XCAD
18	Relational Automatic Statistician	UNIST	Anh Tong, Jaesik Choi	Discovering Latent Covariance Structures for Multiple Time Series	https://github.com/OpenXAIProject/Automatic-Stock-Report
19	SHAP (SHapley Additive exPlanations)	University of Washington	Scott M. Lundberg, Gabriel G. Erion, Su-In Lee	Consistent Individualized Feature Attribution for Tree Ensembles	https://github.com/slundberg/shap

https://github.com/OpenXAIProject


Search or jump to...
Pull requests
Issues
Marketplace
Explore
OpenXAIProject
Repositories 18
People 17
Teams 0
Projects 0
Settings
Search repositories...
Type: All
Language: All
Customize pinned repositories
New
Preprocessed_IXI_Dataset
IXI dataset, preprocessed to normalized brain volume maps.
Updated a day ago
Preprocessed_ABIDE_Dataset
ABIDE Dataset, preprocessed to normalized brain volume maps.
Updated a day ago
LRP-EEG
Classification and Explanation Model of Brain Signal based on Deep Learning Model
Python
Apache-2.0
Updated 8 days ago
xai
Forked from kirarencteon/xai
Top languages
Python
Jupyter Notebook
JavaScript
Perl
Matlab
People
17
Invite someone



OpenXAIProject / PyConKorea2019-Tutorials

Watch 13 | Unstar 37 | Fork 14

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Security | Insights | Settings

Tutorials about Layer-wise Relevance Propagation (LRP) and SHapley Additive exPlanations (SHAP)

Edit

Manage topics

52 commits | 1 branch | 0 releases | 4 contributors | Apache-2.0

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

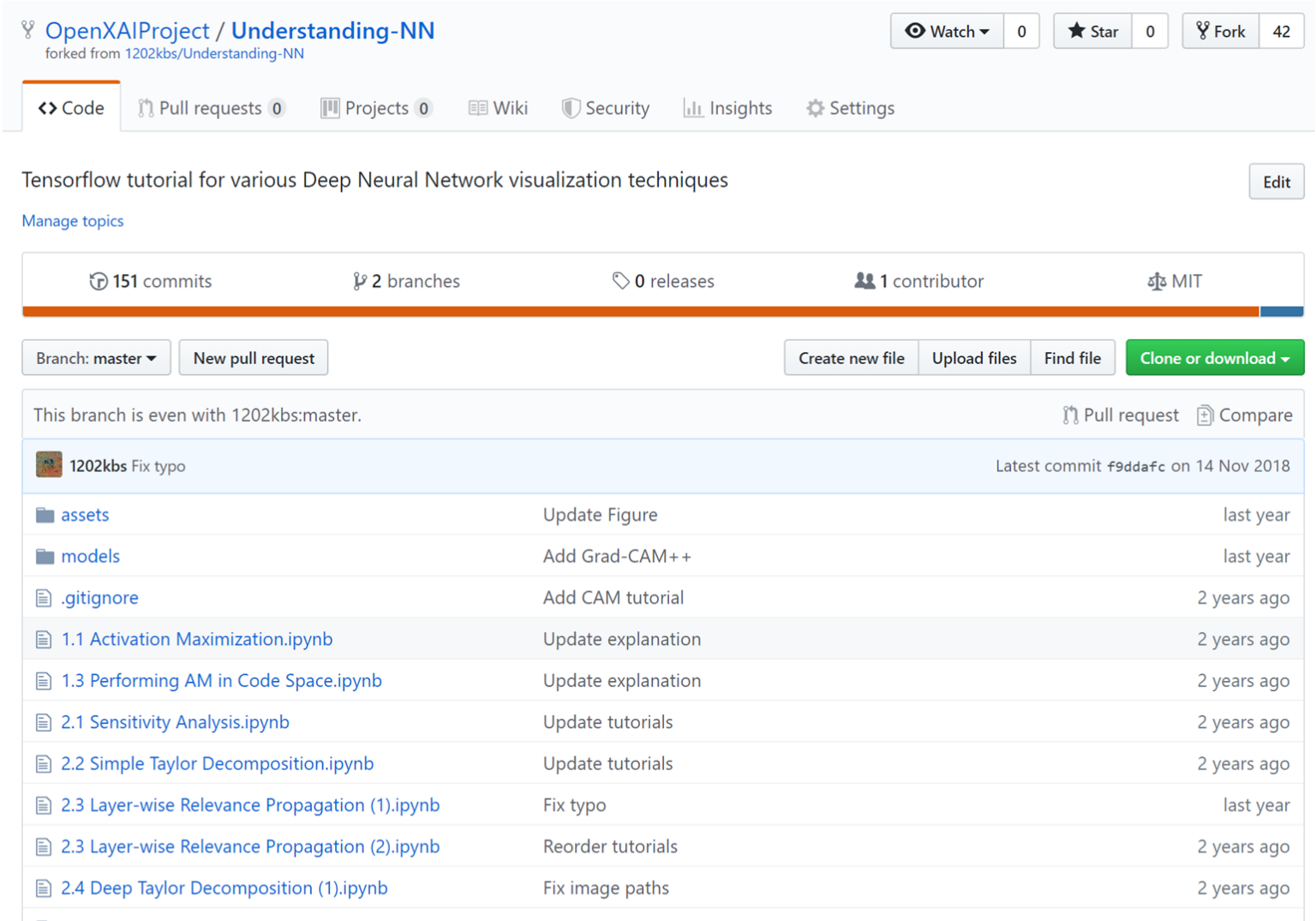
SeongmanHeo Update README.md ... Latest commit 4fa4f78 on 19 Aug

LRP	Add files via upload	3 months ago
SHAP	presentation material	3 months ago
LICENSE	Initial commit	3 months ago
PyConKorea2019-Introduction-presentation.pdf	Add files via upload	3 months ago
README.md	Update README.md	2 months ago
colab_setting.ipynb	Add files via upload	3 months ago

README.md

딥러닝 안에서 일어나는 과정을 설명하는 인공지능 기술 (PyCon Korea 2019 Tutorial)

Introduction



OpenXAIProject / Understanding-NN
forked from 1202kbs/Understanding-NN

Watch 0 | Star 0 | Fork 42

Code | Pull requests 0 | Projects 0 | Wiki | Security | Insights | Settings

Tensorflow tutorial for various Deep Neural Network visualization techniques

Edit

Manage topics

151 commits | 2 branches | 0 releases | 1 contributor | MIT

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

This branch is even with 1202kbs:master. Pull request Compare

1202kbs Fix typo — Latest commit f9ddafc on 14 Nov 2018

assets	Update Figure	last year
models	Add Grad-CAM++	last year
.gitignore	Add CAM tutorial	2 years ago
1.1 Activation Maximization.ipynb	Update explanation	2 years ago
1.3 Performing AM in Code Space.ipynb	Update explanation	2 years ago
2.1 Sensitivity Analysis.ipynb	Update tutorials	2 years ago
2.2 Simple Taylor Decomposition.ipynb	Update tutorials	2 years ago
2.3 Layer-wise Relevance Propagation (1).ipynb	Fix typo	last year
2.3 Layer-wise Relevance Propagation (2).ipynb	Reorder tutorials	2 years ago
2.4 Deep Taylor Decomposition (1).ipynb	Fix image paths	2 years ago
2.4 Deep Taylor Decomposition (2).ipynb	Reorder tutorials	2 years ago
2.5 DeepLIFT.ipynb	Add DeepLIFT tutorial	2 years ago
3.1 Deconvolution.ipynb	Fix typo	2 years ago

https://www.youtube.com/channel/UCGxsfIsOry_LdBaPSet2p7g

YouTube

XAI Open
2 subscribers

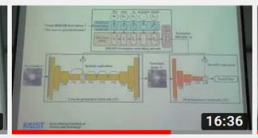
Uploads PLAY ALL SORT BY

28:43

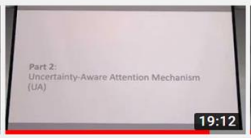
XAI open software project - Few-shot Learning
14 views • 3 weeks ago


17:44

XAI open software project - Training C-C Classifier for...
5 views • 3 weeks ago

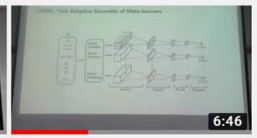

16:36

XAI open software project - Generative model for...
7 views • 3 weeks ago



19:12

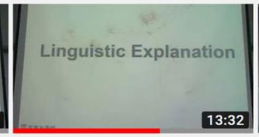
XAI open software project - Implementation Tutorial:...
10 views • 3 weeks ago


6:46

XAI open software project - Task-Adaptive Ensemble of...
3 views • 3 weeks ago

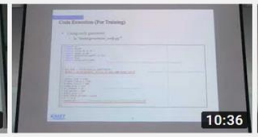
8:57

XAI open software project - Adaptive network...
3 views • 3 weeks ago

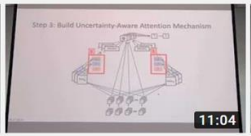


13:32

XAI open software project - Linguistic Explanation
10 views • 3 weeks ago


10:36

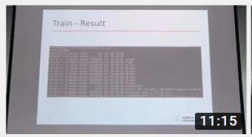
XAI open software project - Subsampling Aliasing...
13 views • 3 weeks ago


11:04

XAI open software project - Implementation Tutorial:...
1 view • 3 weeks ago

8:27

XAI open software project - Data-driven Open-domain...
12 views • 3 weeks ago
CC

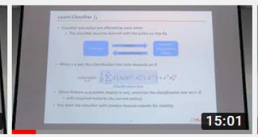


11:15

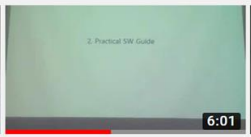
XAI open software project - Data-driven Open-domain...
2 views • 3 weeks ago

6:12

XAI open software project - Data-driven Open-domain...
7 views • 3 weeks ago


15:01

XAI open software project - Joint AFA Classification
9 views • 1 month ago

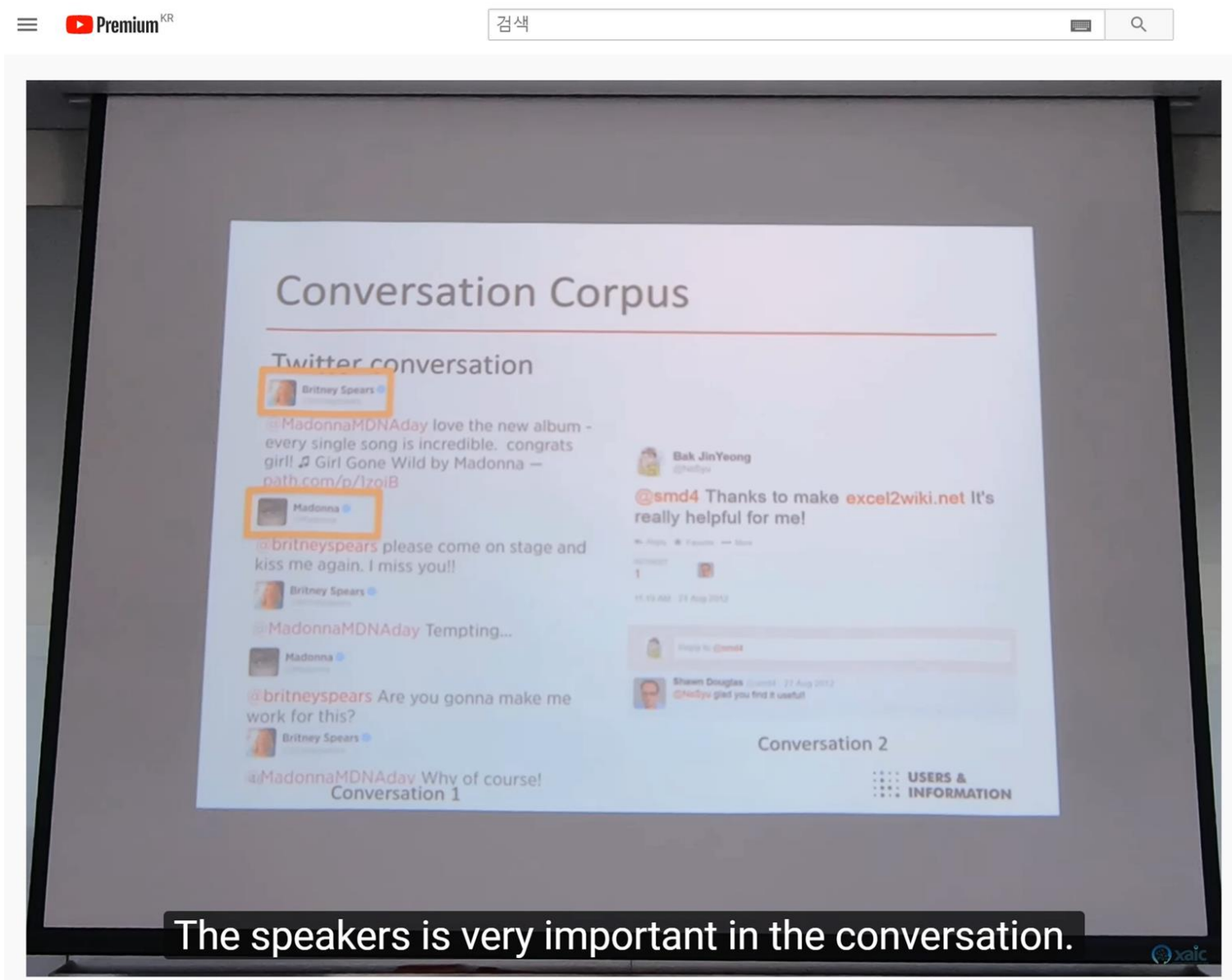


6:01

XAI open software project - Paraphrase Generating...
8 views • 1 month ago

10:51

XAI open software project - Relational Automatic...
13 views • 1 month ago



XAI open software project - Data-driven Open-domain Neural Conversation Models (Part 3)

Thank you!

Let us know other related projects!

sohee.cho@kaist.ac.kr