Giorgos Paraskevopoulos

🌐 **Website:** http://geopar.eu/

Gender: Male   **Date of birth:** 17 Sep 1991   **Nationality:** Greek

ABOUT ME

Giorgos Paraskevopoulos received his Diploma Degree in Electrical Engineering and Computer Science from the National Technical University of Athens in June 2016. For his diploma thesis he developed a gesture control system for a smart room using Kinect data under the supervision of Dr Evaggelos Spyrou in NCSR Demokritos. Since 2017 he is a PhD student in NTUA, under the supervision of Prof. Alexandros Potamianos.He has industry experience both as a software engineer (intracom telecom) and as a machine learning engineer (behavioral signals). In 2019 he worked as an applied research intern with the Alexa team in Amazon Lab126. Since 2020 he is a research collaborator at the Institute for Speech and Language Processing, Athena RC. His research interests revolve around the extraction and fusion of multimodal representations, transfer learning and adaptation of neural networks to unseen domains.

WORK EXPERIENCE

[20 Jul 2020 – Current]   **Associate Researcher**

Institute for Speech and Language Processing   https://www.ilsp.gr/

Address: Artemidos 6, 15125, Maroussi, Greece
City: Athens
Country: Greece
Name of unit or department: ILSP
Business or sector: Professional, scientific and technical activities
Main activities and responsibilities:
  ○ **Safety4All**: The Safety4ALL project aims to engage the customer and trigger his or her involvement in a monitor, report and update loop for the validation and enrichment of safety measures.
  ○ **ASR:** Greek Speech Recognition using Kaldi
  ○ **Theano**: A Greek-speaking conversational agent for COVID-19
  ○ **NLP:** Natural Language Processing tools (spell-corrector, multiword-expressions extractor)

[1 Sep 2017 – Current]   **Higher education teaching assistant**

National Technical University of Athens   https://www.ece.ntua.gr

Address: 9, Iroon Polytechniou St, 15780, Athens, Greece
City: Athens
Country: Greece
Name of unit or department: Electrical and Computer Engineering
Business or sector: Education
Main activities and responsibilities:
Teaching Assistant. Responsible for labs and general assistance for the following courses:
  ○ Pattern Recognition
  ○ Speech and Language Processing
  ○ Signals and System

[1 Aug 2019 – 1 Nov 2019] **Applied Research Intern**

Amazon, inc (Alexa group, Lab 126) https://amazon.jobs/en/teams/lab126/

Address: 1100 Enterprise Way, 94089, Sunnyvale, United States
City: Sunnyvale
Country: United States
Name of unit or department: Alexa, Lab 126
Business or sector: Professional, scientific and technical activities
Main activities and responsibilities:
Applied Research internship, Alexa group, Amazon Lab 126

- Worked on multimodal grounding for Automatic Speech Recognition
- Contributed to internal framework for distributed neural network training
- Paper "Multimodal and Multiresolution Speech Recognition with Transformers" accepted for publica-
tion in ACL2020 conference

[1 Apr 2018 – 30 Apr 2020] **Machine Learning Engineer**

Behavioral Signal Technologies, Inc. https://behavioralsignals.com/

Address: 4470 W Sunset Blvd, Suite , 94260, Los Angeles, United States
City: Athens
Country: Greece
Business or sector: Professional, scientific and technical activities
Main activities and responsibilities:
Machine Learning Engineer
- Improving the performance of Emotion Recognition from speech technology with regard to core KPIs
- Bridging the gap between R&D and production by integrating ML models into CI
- Design and development of baseline NLP models
- Domain adaptation of ML models for different use cases
- Dockerization and performance improvement of deployed models
- Developed an Automatic Speech Recognition system for Greek language
- Quentin Furhatino demo

NOTES: Took a leave of absense between 1 Aug 2019 - 1 Dec 2019 to work as an applied intern in Amazon, Lab 126.

[1 May 2017 – 30 May 2019] **Higher education research assistant**

National Technical University of Athens https://babyrobot.eu/

Address: 9, Iroon Polytechniou St, 15780, Athens, Greece
City: Athens
Country: Greece
Name of unit or department: Electrical and Computer Engineering
Business or sector: Professional, scientific and technical activities
Main activities and responsibilities:
Research assistant, Project website

- Worked on Babyrobot EU project. Babyrobot aims to develop state-of-the-art social robotics with application on ASD children inclusion.
- Developed components for affective processing of children speech
- Integration of project components in a unified micro-service architecture

[1 Sep 2017 – 31 Mar 2018] **Devops Engineer**

Behavioral Signals Technologies, inc   https://behavioralsignals.com/

Address: 4470 W Sunset Blvd, Suite, 94260, Los angeles, United States
City: Athens
Country: Greece
Business or sector: Professional, scientific and technical activities
Main activities and responsibilities:
Responsible for
- Developing a CI/CD pipeline with Jenkins
- Establishing a development process based on Code Reviews and Continuous Integration

[1 Jul 2015 – 30 Apr 2017] **SDN/NFV R&D Engineer**

Intracom Telecom   http://www.intracom-telecom.com/

Address: 19.7 km Markopoulou Ave., 19002, Paiania, Greece
City: Athens
Country: Greece
Name of unit or department: SDN/NFV
Business or sector: Professional, scientific and technical activities
Main activities and responsibilities:
Responsible for

- OpenDaylight SDN controller scalability optimization
- NFV accelerator: Automatic NFV workload optimization
- Contributions to OPNFV open-source project. OPNFV aims to create a reference NFV platform through system level integration, deployment and testing
- Contributing to company driven open-source projects (see https://github.com/intracom-telecom-sdn)

[1 Jan 2015 – 31 Oct 2016] **Research Assistant**

NCSR Demokritos   https://www.facebook.com/synaisthisiproject/

Address: Neapoleos 27 &, Patriarchou Grigoriou, 15310, Aghia Paraskevi, Greece
City: Athens
Country: Greece
Name of unit or department: Institute of Informatics & Telecommunications
Business or sector: Professional, scientific and technical activities
Main activities and responsibilities:
- Developed my thesis project. Aigaio NUI is a gesture based Natural User Interface for a smart room integrated into the SYNAISTHISI platform.
- Technical presentation of the project available on slideshare and on github.
- Demo video available here (the video is in Greek).
- Resulted in the publications A
- A real-time approach for gesture recognition using the Kinect sensor, SETN, 2016
- Real-time arm gesture recognition using 3D skeleton joint data, Algorithms, 2019

EDUCATION AND TRAINING

[1 Sep 2017 – Current] PhD Student

National Technical University of Athens   https://www.ece.ntua.gr/

Address: 9, Iroon Polytechniou St, 15780, Athens, Greece
Field(s) of study: Information and Communication Technologies
Level in EQF: EQF level 8
Thesis: Extraction of deep multimodal representations and application in learning from limited data
Main subject / occupational skills covered:
My PhD thesis is related to my ongoing work on extracting multimodal representations from audio, text and image data. Emphasis is given on techniques involving Transfer learning and Domain adaptation.

[1 Sep 2009 – 1 Jun 2016] Diploma (5-year joint degree)

National Technical University of Athens   https://www.ece.ntua.gr/

Address: 9, Iroon Polytechniou St, 15780, Athens, Greece
Field(s) of study: Electrical & Computer Engineering
Final grade: 7.7   **Level in EQF:** EQF level 7
Type of credits: ECTS   **Number of credits:** 300
Thesis: Smart Room Gesture Control using Kinect skeletal data
Main subject / occupational skills covered:
- Signal Processing, Control & Robotics subjects: (Pattern Recognition, Speech & Language Processing, Computer Vision, Control Systems etc.)
- Computer Science subjects: (Algorithms, Programming Languages, Operating Systems etc.)

[2006 – 2009] High School Diploma

Ellinogermaniki Agogi

Level in EQF: EQF level 3

LANGUAGE SKILLS

Mother tongue(s): Greek

Other language(s):

English
LISTENING C2 **READING** C2 **WRITING** C2
SPOKEN PRODUCTION C2 **SPOKEN INTERACTION** C2

German
LISTENING A1 **READING** A1 **WRITING** A1
SPOKEN PRODUCTION A1 **SPOKEN INTERACTION** A1

DIGITAL SKILLS

Programming Languages
Python | C++ | JavaScript | Go | C | Shell Script (Bash)

Frameworks / Libraries
SpaCy | Tensorflow | Fairseq | Scikit-Learn | Numpy | PyTorch Lightning | Keras | OpenCV | NLTK | PyTorch

Speech Recognition
Kaldi

Tools / OS
Docker | Linux | Git

PROJECTS

[Current] **Kaldi gRPC Server**

This is a modern alternative for deploying Speech Recognition models developed using Kaldi.

https://github.com/georgepar/kaldi-grpc-server

[Current] **slp**

A framework to develop models for natural language and speech processing

https://github.com/georgepar/slp

REVIEWING SERVICE

ACL - 2021 | AAAI - 2021 | ICMI - 2021 | ACM Multimedia - 2021

PUBLICATIONS

[2022]

[1] Paraskevopoulos G., Georgiou E., Potamianos A., MMLatch: Bottom-up Top-down Fusion for Multimodal Sentiment Analysis (2022), to appear in Proceedings of ICASSP 2022
https://arxiv.org/abs/2201.09828
ACCEPTED: ICASSP 2022

Current deep learning approaches for multimodal fusion rely on bottom-up fusion of high and mid-level latent modality representations (late/mid fusion) or low level sensory inputs (early fusion). Models of human perception highlight the importance of top-down fusion, where high-level representations affect the way sensory inputs are perceived, i.e. cognition affects perception. These top-down interactions are not captured in current deep learning models. In this work we propose a neural architecture that captures top-down cross-modal interactions, using a feedback mechanism in the forward pass during network training. The proposed mechanism extracts high-level representations for each modality and uses these representations to mask the sensory inputs, allowing the model to perform top-down feature masking. We apply the proposed model for multimodal sentiment recognition on CMU-MOSEI. Our method shows consistent improvements over the well established MulT and over our strong late fusion baseline, achieving state-of-the-art results.

[2022]

[2] Kouni V., Paraskevopoulos G., Rauhut H., Alexandropoulos GC. "ADMM-DAD net: a deep unfolding network for analysis compressed sensing", to appear in Proceedings of ICASSP 2022

https://arxiv.org/abs/2110.06986

ACCEPTED ICASSP 2022

In this paper, we propose a new deep unfolding neural network based on the ADMM algorithm for analysis Compressed Sensing. The proposed network jointly learns a redundant analysis operator for sparsification and reconstructs the signal of interest. We compare our proposed network with a state-of-the-art unfolded ISTA decoder, that also learns an orthogonal sparsifier. Moreover, we consider not only image, but also speech datasets as test examples. Computational experiments demonstrate that our proposed network outperforms the state-of-the-art deep unfolding networks, consistently for both real-world image and speech datasets.

[2021]

[3] Melistas T., Giannakopoulos T., Paraskevopoulos G. "Lyrics and Vocal Melody Generation conditioned on Accompaniment", In Proceedings of the 2nd Workshop on NLP for Music and Audio (NLP4MusA 2021)

https://sites.google.com/view/nlp4musa-2021/accepted-papers

In this paper we present a previously unexplored task, the generation of lyrics and vocal melody for a given instrumental music piece in the symbolic domain. We model the above as a sequence-to-sequence task, using a memory efficient Transformer architecture, which we train on text event sequences that describe entire songs. Towards this end, we build a suitable dataset and apply musical analysis, compressing the instrumental part and making it key-independent. We further design a novel architecture to decouple lyrics and melody generation, making it possible to use pretrained language models and conditioning on lyrics. Finally, Mellotron is used to turn the generated sequences into singing audio.

[2021]

[4] Georgiou E., Paraskevopoulos G., Potamianos A. "M3: MultiModal Masking applied to sentiment analysis", Interspeech 2021

https://www.isca-speech.org/archive/interspeech_2021/georgiou21_interspeech.html

A common issue when training multimodal architectures is that not all modalities contribute equally to the model's prediction and the network tends to over-rely on the strongest modality. In this work, we present M3, a training procedure based on modality masking for deep multimodal architectures. During network training, we randomly select one modality and mask its features, forcing the model to make its prediction in the absence of this modality. This structured regularization allows the network to better exploit complementary information in input modalities. We implement M3 as a generic layer that can be integrated with any multimodal architecture. Our experiments show that M3 outperforms other masking schemes and improves performance for our strong baseline. We evaluate M3 for multimodal sentiment analysis on CMU-MOSEI, achieving results comparable to the state-of-the-art.

[2021]

[5] Zaranis E., Paraskevopoulos G., Katsamanis A., Potamianos A., "EmpBot: A T5-based Empathetic Chatbot focusing on Sentiments", CoRR preprint, (2021)

https://arxiv.org/abs/2111.00310

In this paper, we introduce EmpBot: an end-to-end empathetic chatbot. Empathetic conversational agents should not only understand what is being discussed, but also acknowledge the implied feelings of the conversation partner and respond appropriately. To this end, we propose a method based on a transformer pretrained language model (T5). Specifically, during finetuning we propose to use three objectives: response language modeling, sentiment understanding, and empathy forcing. The first objective is crucial for generating relevant and coherent responses, while the next ones are significant for acknowledging the sentimental state of the conversational partner and for favoring empathetic responses. We evaluate our model on the EmpatheticDialogues dataset using both automated metrics and human evaluation. The inclusion of the sentiment understanding and empathy forcing auxiliary losses favor empathetic responses, as human evaluation results indicate, comparing with the current state-of-the-art.

[2021]

[6] Ventoura N., Vassilakis Y., Palios K., Paraskevopoulos G., Katsamanis A., Katsouros V. "Theano: A Greek-speaking conversational agent for COVID-19", In Proceedings of the 1st Workshop on NLP for Positive Impact, 2021

https://aclanthology.org/2021.nlp4posimpact-1.5/

Conversational Agents (CAs) can be a proxy for disseminating information and providing support to the public, especially in times of crisis. CAs can scale to reach larger numbers of end-users than human operators, while they can offer information interactively and engagingly. In this work, we present Theano, a Greek-speaking virtual assistant for COVID-19. Theano presents users with COVID-19 statistics and facts and informs users about the best health practices as well as the latest COVID-19 related guidelines. Additionally, Theano provides support to end-users by helping them self-assess their symptoms and redirecting them to first-line health workers. The relevant, localized information that Theano provides, makes it a valuable tool for combating COVID-19 in Greece. Theano has already conversed with different users in more than 170 different conversations through a web interface as a chatbot and over the phone as a voice bot.

[2021]

[7] Karouzos C., Paraskevopoulos G., Potamianos A. "UDALM: Unsupervised Domain Adaptation through Language Modeling", Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

https://aclanthology.org/2021.naacl-main.203

In this work we explore Unsupervised Domain Adaptation (UDA) of pretrained language models for downstream tasks. We introduce UDALM, a fine-tuning procedure, using a mixed classification and Masked Language Model loss, that can adapt to the target domain distribution in a robust and sample efficient manner. Our experiments show that performance of models trained with the mixed loss scales with the amount of available target data and the mixed loss can be effectively used as a stopping criterion during UDA training. Furthermore, we discuss the relationship between A-distance and the target error and explore some limitations of the Domain Adversarial Training approach. Our method is evaluated on twelve domain pairs of the Amazon Reviews Sentiment dataset, yielding 91.74% accuracy, which is an 1.11% absolute improvement over the state-of-the-art.

[2020]

[8] Xezonaki D., Paraskevopoulos G., Potamianos A., Narayanan S. "Affective Conditioning on Hierarchical Attention Networks applied to Depression Detection from Transcribed Clinical Interviews", Interspeech 2020

https://www.isca-speech.org/archive/interspeech_2020/xezonaki20_interspeech.html

In this work we propose a machine learning model for depression detection from transcribed clinical interviews. Depression is a mental disorder that impacts not only the subject's mood but also the use of language. To this end we use a Hierarchical Attention Network to classify interviews of depressed subjects. We augment the attention layer of our model with a conditioning mechanism on linguistic features, extracted from affective lexica. Our analysis shows that individuals diagnosed with depression use affective language to a greater extent than not-depressed. Our experiments show that external affective information improves the performance of the proposed architecture in the General Psychotherapy Corpus and the DAIC-WoZ 2017 depression datasets, achieving state-of-the-art 71.6 and 70.3 using the test set, F1-scores respectively.

[2020]

[9] Paraskevopoulos G., Parthasarathy S., Khare A., Sundaram S. "Multimodal and Multiresolution Speech Recognition with Transformers". In Proceedings of the 58th Annual Meeting of the Association for Compu- tational Linguistics (2020)

https://aclanthology.org/2020.acl-main.216/

This paper presents an audio visual automatic speech recognition (AV-ASR) system using a Transformer-based architecture. We particularly focus on the scene context provided by the visual information, to ground the ASR. We extract representations for audio features in the encoder layers of the transformer and fuse video features using an additional crossmodal multihead attention layer. Additionally, we incorporate a multitask training criterion for multiresolution ASR, where we train the model to generate both character and subword level transcriptions. Experimental results on the How2 dataset, indicate that multiresolution training can speed up convergence by around 50% and relatively improves word error rate (WER) performance by upto 18% over subword prediction models. Further, incorporating visual information improves performance with relative gains upto 3.76% over audio only models. Our results are comparable to state-of-the-art Listen, Attend and Spell-based architectures.

[2019]

[10] Paraskevopoulos G., Spyrou E., Sgouropoulos D., Giannakopoulos T., Mylonas P. "Real-time arm gesture recognition using 3D skeleton joint data". Algorithms, 12(5), p.108 (2019)

https://www.mdpi.com/1999-4893/12/5/108

In this paper we present an approach towards real-time hand gesture recognition using the Kinect sensor, investigating several machine learning techniques. We propose a novel approach for feature extraction, using measurements on joints of the extracted skeletons. The proposed features extract angles and displacements of skeleton joints, as the latter move into a 3D space. We define a set of gestures and construct a real-life data set. We train gesture classifiers under the assumptions that they shall be applied and evaluated to both known and unknown users. Experimental results with 11 classification approaches prove the effectiveness and the potential of our approach both with the proposed dataset and also compared to state-of-the-art research works.

[2019]

[11] Chatziagapi A., Paraskevopoulos G., Sgouropoulos D., Pantazopoulos G., Nikandrou M., Giannakopoulos T., Katsamanis A., Potamianos A., Narayanan S. "Data Augmentation using GANs for Speech Emotion Recognition". Proc. Interspeech 2019 (pp. 171-175)

https://www.isca-speech.org/archive_v0/Interspeech_2019/abstracts/2561.html

In this work, we address the problem of data imbalance for the task of Speech Emotion Recognition (SER). We investigate conditioned data augmentation using Generative Adversarial Networks (GANs), in order to generate samples for underrepresented emotions. We adapt and improve a conditional GAN architecture to generate synthetic spectrograms for the minority class. For comparison purposes, we implement a series of signal-based data augmentation methods. The proposed GAN-based approach is evaluated on two datasets, namely IEMOCAP and FEEL-25k, a large multi-domain dataset. Results demonstrate a 10% relative performance improvement in IEMOCAP and 5% in FEEL-25k, when augmenting the minority classes.

[2019]

[12] Paraskevopoulos, G., Tzinis, E., Ellinas, N., Giannakopoulos, T., Potamianos, A. "Unsupervised low-rank representations for speech emotion recognition". In Proceedings of Interspeech 2019 (pp. 939-943)

https://www.isca-speech.org/archive/interspeech_2019/paraskevopoulos19_interspeech.html

We examine the use of linear and non-linear dimensionality reduction algorithms for extracting low-rank feature representations for speech emotion recognition. Two feature sets are used, one based on low-level descriptors and their aggregations (IS10) and one modeling recurrence dynamics of speech (RQA), as well as their fusion. We report speech emotion recognition (SER) results for learned representations on two databases using different classification methods. Classification with low-dimensional representations yields performance improvement in a variety of settings. This indicates that dimensionality reduction is an effective way to combat the curse of dimensionality for SER. Visualization of features in two dimensions provides insight into discriminatory abilities of reduced feature sets.

[2018]

[13] Paraskevopoulos, G., Tzinis, E., Vlatakis, E., Potamianos, A. "Pattern Search Multidimensional Scaling", CoRR preprint, (2018)

https://arxiv.org/abs/1806.00416

We present a novel view of nonlinear manifold learning using derivative-free optimization techniques. Specifically, we propose an extension of the classical multi-dimensional scaling (MDS) method, where instead of performing gradient descent, we sample and evaluate possible "moves" in a sphere of fixed radius for each point in the embedded space. A fixed-point convergence guarantee can be shown by formulating the proposed algorithm as an instance of General Pattern Search (GPS) framework. Evaluation on both clean and noisy synthetic datasets shows that pattern search MDS can accurately infer the intrinsic geometry of manifolds embedded in high-dimensional spaces. Additionally, experiments on real data, even under noisy conditions, demonstrate that the proposed pattern search MDS yields state-of-the-art results.

[2018]

[14] Tzinis, E., Paraskevopoulos, G., Baziotis, C., and Potamianos, A. "Integrating Recurrence Dynamics for Speech Emotion Recognition". In Proceedings of Interspeech 2018 (pp. 927931)

https://www.isca-speech.org/archive/interspeech_2018/tzinis18_interspeech.html

We investigate the performance of features that can capture nonlinear recurrence dynamics embedded in the speech signal for the task of Speech Emotion Recognition (SER). Reconstruction of the phase space of each speech frame and the computation of its respective Recurrence Plot (RP) reveals complex structures which can be measured by performing Recurrence Quantification Analysis (RQA). These measures are aggregated by using statistical functionals over segment and utterance periods. We report SER results for the proposed feature set on three databases using different classification methods. When fusing the proposed features with traditional feature sets, e.g., [1], we show an improvement in unweighted accuracy of up to 5.7% and 10.7% on Speaker-Dependent (SD) and Speaker-Independent (SI) SER tasks, respectively, over the baseline [1]. Following a segment-based approach we demonstrate state-of-the-art performance on IEMOCAP using a Bidirectional Recurrent Neural Network.

[2018]

[15] Baziotis, C., Nikolaos, A., Chronopoulou, A., Kolovou, A., Paraskevopoulos, G., Ellinas, N., Narayanan, S., and Potamianos, A. "NTUA-SLP at SemEval-2018 Task 1: Predicting Affective Content in Tweets with Deep Attentive RNNs and Transfer Learning". In Proceedings of The 12th International Workshop on Semantic Evaluation (pp. 245255). New Orleans, Louisiana NAACL, 2018

https://aclanthology.org/S18-1037/

In this paper we present deep-learning models that submitted to the SemEval-2018 Task 1 competition: "Affect in Tweets". We participated in all subtasks for English tweets. We propose a Bi-LSTM architecture equipped with a multi-layer self attention mechanism. The attention mechanism improves the model performance and allows us to identify salient words in tweets, as well as gain insight into the models making them more interpretable. Our model utilizes a set of word2vec word embeddings trained on a large collection of 550 million Twitter messages, augmented by a set of word affective features. Due to the limited amount of task-specific training data, we opted for a transfer learning approach by pretraining the Bi-LSTMs on the dataset of Semeval 2017, Task 4A. The proposed approach ranked 1st in Subtask E "Multi-Label Emotion Classification", 2nd in Subtask A "Emotion Intensity Regression" and achieved competitive results in other subtasks.

[2018]

[16] Baziotis, C., Nikolaos, A., Kolovou, A., Paraskevopoulos, G., Ellinas, N., and Potamianos, A. "NTUA- SLP at SemEval-2018 Task 2: Predicting Emojis using RNNs with Context-aware Attention". In Proceedings of The 12th International Workshop on Semantic Evaluation (pp. 438444). New Orleans, Louisiana NAACL, 2018

https://aclanthology.org/S18-1069/

In this paper we present a deep-learning model that competed at SemEval-2018 Task 2 "Multilingual Emoji Prediction". We participated in subtask A, in which we are called to predict the most likely associated emoji in English tweets. The proposed architecture relies on a Long Short-Term Memory network, augmented with an attention mechanism, that conditions the weight of each word, on a "context vector" which is taken as the aggregation of a tweet's meaning. Moreover, we initialize the embedding layer of our model, with word2vec word embeddings, pretrained on a dataset of 550 million English tweets. Finally, our model does not rely on hand-crafted features or lexicons and is trained end-to-end with back-propagation. We ranked 2nd out of 48 teams.

[2018]

[17] Baziotis, C., Nikolaos, A., Papalampidi, P., Kolovou, A., Paraskevopoulos, G., Ellinas, N., and Potami- anos, A, "NTUA-SLP at SemEval-2018 Task 3: Tracking Ironic Tweets using Ensembles of Word and Char- acter Level Attentive RNNs". In Proceedings of The 12th International Workshop on Semantic Evaluation (pp. 613-621). New Orleans, Louisiana NAACL, 2018

https://aclanthology.org/S18-1100/

In this paper we present two deep-learning systems that competed at SemEval-2018 Task 3 "Irony detection in English tweets". We design and ensemble two independent models, based on recurrent neural networks (Bi-LSTM), which operate at the word and character level, in order to capture both the semantic and syntactic information in tweets. Our models are augmented with a self-attention mechanism, in order to identify the most informative words. The embedding layer of our word-level model is initialized with word2vec word embeddings, pretrained on a collection of 550 million English tweets. We did not utilize any handcrafted features, lexicons or external datasets as prior information and our models are trained end-to-end using back propagation on constrained data. Furthermore, we provide visualizations of tweets with annotations for the salient tokens of the attention layer that can help to interpret the inner workings of the proposed models. We ranked 2nd out of 42 teams in Subtask A and 2nd out of 31 teams in Subtask B. However, post-task-completion enhancements of our models achieve state-of-the-art results ranking 1st for both subtasks.

[2017]

[18] Paraskevopoulos, G., Karamanolakis, G., Iosif, E., Pikrakis, A., Potamianos, A., "Sensory-Aware Multi- modal Fusion for Word Semantic Similarity Estimation". MultiLearn2017: Multimodal Processing, Modeling and Learning for Human-Computer/Robot Interaction Workshop, in conjuction with EUSIPCO-2017

https://www.eurasip.org/Proceedings/Eusipco/Eusipco2017/wpapers/ML4.pdf

Traditional semantic models are disembodied from the human perception and action. In this work, we attempt to address this problem by grounding semantic representations of words to the acoustic and visual modalities. Specifically we estimate multimodal word representations via the fusion of auditory and visual modalities with the text modality. We employ middle and late fusion of representations with modality weights assigned to each of the unimodal representations. We also propose a fusion method that assigns different weights to each word, based on how relevant that word is for the audio and visual modalities. The proposed methods are evaluated for the task of semantic similarity computation between words. To our knowledge, this is the first work that combines text, audio and visual features for the computation of multimodal semantic word representations. Multimodal models outperform the unimodal models, indicating the importance of multimodal fusion and perceptual grounding.

[2016]

[19] Paraskevopoulos, G., Spyrou, E., Sgouropoulos, D., "A Real-Time Approach for Gesture Recognition Using the Kinect Sensor". In Proceedings of the 9th Hellenic Conference on Artificial Intelligence (pp. 31:1–31:4), Thessaloniki, Greece, SETN 2016.

https://dl.acm.org/doi/10.1145/2903220.2903241

In this paper we present an approach for real-time gesture recognition using the Kinect sensor and a set of machine learning techniques. We propose a novel approach for feature extraction using measurements on skeletal joints. We select a set of simple gestures and construct a data set. We train classifiers under the assumptions that they shall be evaluated to both known and unknown users. Experimental results prove the effectiveness and the potential of our approach.