Chieh-Yang Huang 黃介揚

Ph.D. Student
College of Information Sciences and Technology
Penn State University

Email: chiehyang@psu.edu
appleternity@gmail.com
Website: http://chieh-yang.co

Education

- **PhD. Student, Informatics**
  College of IST, Penn State University, Jan 2019 – Dec 2022 (anticipated)
- **PhD. Student, Computer Science (transferred to PSU)**
  School of CIDSE, Arizona State University, Aug 2017 – Dec 2018
- **B.S., Electrical Engineering**
  Dept. Electrical Engineering, National Taiwan University, Sep 2010 – June 2014

Research Interest

My research interest lies in the intersection of **Natural Language Processing (NLP)** and **Human Computer Interaction (HCI)**, where I build AI-powered and crowd-powered systems to support writing, language learning, and other language-related tasks.

Papers

[P.17] Anton Belyy*, **Chieh-Yang Huang** *, Jacob Andreas, Emmanouil Antonios Platanios, Sam Thomson, Richard Shin, Subhro Roy, Aleksandr Nisnevich, Charles Chen, Benjamin Van Durme (2022). Guided K-best Selection for Semantic Parsing Annotation. To appear in the 60th Annual Meeting of the Association for Computational Linguistics - Demo Track (ACL 2022 Demo).
*Equal contribution

[P.16] **Chieh-Yang Huang**, Jinfeng Li, Nikita Bhutani, Alexander Whedon, Estevam Hruschka, Yoshihiko Suhara (2022). Extracting Salient Facts from Company Reviews with Scarce Labels. To appear in the 5th instalment of the Fact Extraction and VERification FEVER 2022 (FEVER 2022) co-located with ACL 2022.

[P.15] Jinfeng Li, Nikita Bhutani, Alexander Whedon, **Chieh-Yang Huang**, Estevam Hruschka, Yoshihiko Suhara (2022). Extracting Salient Facts from Company Reviews with Scarce Labels. To appear in the AAAI 2022 Workshop on Artificial Intelligence with Biased or Scarce Data (AIBSD 2022).

[P.14] **Chieh-Yang Huang** and Ting-Hao K. Huang (2021). Semantic Frame Forecasting. In the Proceedings of the 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL 2021**), 6 June - 11 June, 2021, Virtual Conference.

[P.13] Chacha Chen, **Chieh-Yang Huang**, Yaqi Hou, Yang Shi, Enyan Dai, Jiaqi Wang. (2020). TEST_POSITIVE at W-NUT 2020 Shared Task-3: Cross-task modeling. In Proceedings of the Sixth Workshop on Noisy User-generated Text (**W-NUT 2020**), 16 November - 20 November, Virtual Conference.

[P.12] Yun-Hsuan Jen, **Chieh-Yang Huang**, Mei-Hua Chen, Ting-Hao K. Huang, Lun-Wei Ku. (2020). Assessing the Helpfulness of Learning Materials with Inference-Based Learner-Like Agent. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2020**), 16 November - 20 November, Virtual Conference.

[P.11] Ting-Hao K. Huang, **Chieh-Yang Huang**, Chieh-Kuang Cornelia Ding, Yen-Chia Hsu, C Lee Giles. (2020). CODA-19: Reliably Annotating Research Aspects on 10,000+ CORD-19 Abstracts Using a Non-Expert Crowd. In Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020, 5 July - 10 July, 2020, Virtual Conference.

[P.10] **Chieh-Yang Huang**, Shih-Hong Huang, and Ting-Hao K. Huang. (2020). Heteroglossia: In-Situ Story Ideation with the Crowd. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (**CHI 2020**), 25 April - 30 April, 2020, Honolulu, USA.

[P.9] **Chieh-Yang Huang**, Yi-Ting Huang, Mei-Hua Chen and Lun-Wei Ku. (2019). From Receptive to Productive: Learning to Use Confusing Words through Automatically Selected Example Sentences. In the 14th Workshop on Innovative Use of NLP for Building Educational Applications (**BEA 2019**), 28 July - 2 August, 2019, Florence, Italy.

[P.8] Ting-Yao Hsu, **Chieh-Yang Huang**, Yen-Chia Hsu, and Ting-Hao K. Huang. (2019). Visual Story Post-Editing. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. (**ACL 2019**), 28 July - 2 August, 2019, Florence, Italy.

[P.7] **Chieh-Yang Huang**, Hanghang Tong, Jingrui He, and Ross Maciejewski. (2019). Location Prediction for Tweets. Frontiers in Big Data.

[P.6] **Chieh-Yang Huang** and Lun-Wei Ku. (2018). EmotionPush: Emotion and Response Time Prediction towards Human-like Chatbots. In Proceedings of the 2018 of the IEEE Global Communications Conference (**GLOBECOM 2018**), 9-13 December, 2018, Abu Dhabi, UAE.

[P.5] **Chieh-Yang Huang**, Tristan Labetoulle, Ting-Hao K. Huang, Yi-Pei Chen, Hung-Chen Chen, Vallari Srivastava, and Lun-Wei Ku. (2017). MoodSwipe: A Soft Keyboard that Suggests Messages Based on User-Specified Emotions. In the Demo track of the Conference on Empirical Methods in Natural Language Processing 2017 (**EMNLP Demo 2017**), 7–11 September, 2017, Copenhagen, Denmark.

[P.4] **Chieh-Yang Huang**, Mei-Hua, and Lun-Wei Ku (2018). Towards a Better Learning of Near-Synonyms: Automatically Suggesting Example Sentences via Filling in the Blank. In Proceedings of the 26th International Conference on World Wide Web Companion. (**WWW 2017**), 3-7 April, 2017, Perth, Australia.

[P.3] **Chieh-Yang Huang**, Ting-Hao K. Huang, and Lun-Wei Ku. (2017). Challenges in Providing Automatic Affective Feedback in Instant Messaging Applications. In the Designing the User Experience of Machine Learning Systems symposium **(AAAI 2017 Spring Symposium Series)**, 27-29 March, 2017, Palo Alto, USA.

[P.2] **Chieh-Yang Huang**, Nicole Peinelt, and Lun-Wei Ku. (2016). Automatically Suggesting Example Sentences of Near-Synonyms for Language Learners. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations. (**COLING Demo 2016**), 13-16 December, 2016, Osaka, Japan.

[P.1] **Chieh-Yang Huang** and Lun-Wei Ku. (2016). GiveMeExample: Learning confusing words by example sentences. In the Demo Track of the Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. (**ASONAM Demo 2016**), 18-21 August, 2016, San Francisco, USA.

Research Experience and Projects

- **Penn State University**, State College, PA, USA, 01/2019 – present.
  Research Assistant, Advisor: Dr. Ting-Hao K. Huang.
  Research Focuses: Crowdsourcing, NLP, Deep Learning, HCI, Writing Support System.

  – **Automated Story Plot Ideation in Different Forms**
  We explore two ways to convey story plot ideas to writers – word-clouds and plot generation. Word-clouds illustrate possible conceptual words for the follow-up story; and Plot generation produces follow-up plot ideas in natural language. We analyze how people react to and use the two provided forms of ideation.

  – **Forthcoming Semantic Frame Prediction**
  Automatic story plot generation can be hard, especially for long stories. In this task, we model a long story as a sequence of story blocks and represent each story block as a fixed-length of TF-IDF vector over semantic frames. The proposed frame representation could help us capture information for longer stories (more than 50 sentences) and is human understandable. [NAACL 2021]

  – **Mathify: Equation Generation for Novice Researchers**
  Writing equations is a way for researchers to convey ideas without ambiguities but the conventions of writing equations are not clear for novices. This task aims at helping novices to write better equations by automatically generating equations using the foregoing paragraph.

  – **Heteroglossia: Modeling and Supporting Creative Writing**
  Heteroglossia aims at supporting creative writers by ideation. Creative writers can retrieve new ideas for their story drafts from Heteroglossia. We currently provide a "role-playing" strategy to stimulate crowd workers to generate ideas. A Mturk experiment and a deployment study have been conducted and the result shows that Heteroglossia can help creative writers in various aspects. [CHI 2020]

  – **Visual Story Post-Editing.**
  Editing plays an important role for humans. Needless to say, machines need post-editing. In this task, we introduce the first Visual Story Telling Post-Editing dataset, VIST-Edit, and show that post-editing can improve the story quality. The result also suggests that new auto evaluation metrics are needed due to the low correlation between the human judgments and the existing auto metrics. [ACL 2019]

  – **Learner-Like Agent.**
  We assume that learners will perform better when learning from good materials, and vice versa. Given this we build a model that mimics learner's behavior and utilize such a behavior to select good materials automatically. [EMNLP2020]

  – **SlowJimmy.**
  Voice assistants have been researched for decays but have not yet been smart enough to handle tasks spanning in our day-to-day life. We introduce SlowJimmy, a crowd-powered system embedded on Amazon Echo devices. Unlike the previous text-based crowd-powered conversational system, the unbalanced requirements from workers (text) and users (voice) can cause severe problems. In this task, we deploy SlowJimmy and investigate the potential issues.

  – **CaptionThis.**
  Image captioning system can only be applied on the whole image to generate the overview information. CaptionThis aims at exploring the possibility of a "navigable" image caption system, which may help visually impaired people have a better understanding of the image.

- **Semantic Machines - Microsoft Research**, Berkeley, CA, USA, 05/2021 – 08/2021.
  Research Intern, Mentors: Charles Chen, Dr. Jacob Andreas, and Dr. Ben Van Durme.
  Research Focuses: HCI, NLP, AI-powered system.

  - **Computer Assisted Parsing**
    Annotating semantic representation data with difficult schemes is hard. In this task, we explore different AI-powered methods to support annotators. We show that the writing-support approaches can also help in this case.
    [ACL Demo 2022]

- **Megagon Labs**, Mountain View, CA, USA, 06/2020 – 08/2020.
  Research Intern, Mentors: Dr. Jinfeng Li, Dr. Nikita Bhutani, Alexander Whedon, Dr. Yoshi Suhara, and Dr. Wang-Chiew Tan.
  Research Focuses: NLP, Deep Learning.

  - **Automatically Salient Facts Extraction**
    Reading reviews is an important way for people to gather information but a huge portion of reviews are subjective, long, and uninformative. In this task, we define salient facts to be sentences that contain unique, uncommon, and objective information toward an entity and propose several approaches to automatically extract salient facts from a large amount of reviews.
    [AIBSD 2022] [FEVER 2022]

- **Arizona State University**, Tempe, AZ, USA, 05/2018 – 12/2018.
  Research Assistant, Advisor: Dr. Hanghang Tong.
  Research Focuses: Text Mining, Deep Learning.

  - **Geographic Information Prediction on Twitter**
    Geographic Information plays an important role on both marketing and event mining, but is usually blocked due to the privacy issues. This project introduces a deep learning architecture taking the attention mechanism, the subword feature, and the location hierarchy structure into account to predict the geographic information for a given post on Twitter.
    [Frontiers in Big Data.]

- **Institute of Information Science, Academia Sinica**, Taipei, Taiwan, 04/2015 – 06/2017.
  Research Assistant, Advisor: Dr. Lun-Wei Ku.
  Research Focuses: Computer-Assisted Language Learning, Deep Learning, Computer-Mediated Communication, Emotion Detection, Sentiment Analysis.

  - **GiveMeExample: Learning Synonyms by Example Sentences**
    GiveMeExample aims to suggest critical example sentences for language learner to clarify the confusion of synonym. Three main components, the sentence difficulty assessment built by a regression model, the word-sentence fitness estimator built by GMM and BiLSTM, and the heuristic clarification scoring function are introduced to solve this problem. Several websites are built for collecting data and holding evaluation tests.
    [BEA 2019] [WWW 2017] [COLING Demo 2016] [ASONAM Demo 2016]
  - **EmotionPush: Color-Based Emotion Cues for Messaging Applications**
    EmotionPush provides a machine-learning-powered system that automatically conveys users' emotions in messages by color-based emotion cues to bridge the limitation of text-based chatting system in expressing rich emotion.
    [AAAI Spring Symposia 2017]

  - **MoodSwipe: A Keyboard for Sentence Suggestion According to Emotions.**
    MoodSwipe is a mobile phone keyboard that suggests text messages according to the user-specified emotion. We aim to create a convenient user interface to enjoy the technology of emotion classification and text suggestion, and at the same time to collect labeled data automatically. Two emotion classifier models, CNN and LSTM, and two sentence suggestion models, BM25 and similarity of sentence embedding, are built for MoodSwipe.
    [EMNLP Demo 2017]
  - **Response Time Prediction**
    This project aims to predict the response time of a given message sending on the instance message system. This task could be viewed as a measurement of the dialog generation system. A deep learning model integrating conversation and some user-specific information is proposed.
    [GLOBECOM 2018]

- **Communication & Multimedia Lab, National Taiwan University**, Taipei, Taiwan, 09/2013 – 06/2014.
  Undergraduate Intern Student, Advisor: Dr. Yung-Yu Chuang.
  Research Focuses: Digital Image Processing, Machine Learning.

  - **Light Field Image Multi-Label Assignment using Graph Cut.**
    The depth information is a unique feature of light field images. In this project, we aim to integrate the depth information into the multi-label assignment problem without calculating the depth explicitly. A light field image is then treated as a four dimensional image and a revised four dimensional graph cut algorithm is applied to solve this problem.
  - **Shape-Preserving As-Projective-As-Possible (APAP) Image Stitching.**
    APAP algorithm usually causes a distortion when stitching many images. Therefore, we aim to solve this problem by introducing an energy function containing deformation term, line preserving term, and APAP term on the mesh warping algorithm.

- **Speech Processing Lab, National Taiwan University**, Taipei, Taiwan, 09/2012 – 06/2014.
  Undergraduate Intern Student, Advisor: Dr. Lin-Shan Lee.
  Research Focuses: Speech Processing, Machine Learning.

  - **A Dialogue Game Framework with Personalized Training Using Reinforcement Learning.**
    This project aims to help language learners practice speaking through a dialogue game framework. A reinforcement learning agent is proposed to select the path of the dialogue tree in order to maximize the learning efficiency. To improve the pronunciation scoring system, we further integrated phonological features to a neural network model.

Activity and Leadership Experience

- **Director, Senior Yearbook Editor**
  Dept. Electrical Engineering, Taipei, Taiwan, 09/2013 – 06/2014

  - Familiar with Photoshop, Illustrator, and graphic design skills.
  - Leaded a team of ten people to edit the yearbook.

- **Director, Department of Art and Design**
  Electrical Engineering Student Association, Taipei, Taiwan, 09/2012 – 06/2013

  - Leaded a ten-person team to do graphic design and build installed artworks.
  - Supported activities such as Electrical Engineering Night, Christmas Prom, Freshmen Orientation Camp, Electrical Engineering Camp and so on.

Service and Teaching

- **Teaching Assistant**: IST261 - Application Development Design Studio I (Sprint 2022), DS440 - Capstone (Fall 2021), DS440 - Capstone (Spring 2021), IST 402 - Crowdsourcing & Crowd-AI Systems (Spring 2020)
- **Reviewer**: ACL Rolling Review, In2Writing Workshop 2022, UIST 2022, AAAI 2022/2021, EMNLP 2021, ACL 2021, NAACL 2021, CHI LBW 2021, WNUT 2020

References

- **Lun-Wei Ku**
  Associate Research Fellow, Institute of Information Science, Academia Sinica, Taiwan.
  lwku@iis.sinica.edu.tw
- **Ting-Hao K. Huang**
  Tenure-Track Assistant Professor, College of Information Sciences and Technology, Pennsylvania State University, USA
  txh710@psu.edu

Last updated: April 21, 2022