COMP 6721 Applied Artificial Intelligence (Fall 2021)

Project Assignment, Part II

Due date (Moodle Submission): Tuesday, December 7
Counts for 50% of the course project

AI Face Mask Detector: Update. The goal of this second part of the project is to fix any issues that came up during the demo of the first part. Additionally, you have to perform an evaluation for a possible *bias* of your AI and (at least partially) eliminate it.

Bias in AI. With an increasing number of AI-based applications being deployed in practice, the analysis of a possible *bias* – introduced through the training data – has become a major concern.[1] In this project part, your goal is to analyze your AI for **two** of the following categories:[2] *age, race,* or *gender*. For example, based on suitably annotated testing data, you can check if your performance for the four classes (cloth mask/surgical mask/FFP2 mask/no mask) is consistent for male and female faces.[3] To eliminate the bias, experiment with re-balancing and enhancing your training dataset. Evaluate the performance of your network both on the complete dataset and the chosen bias attribute subset(s).

Evaluation: K-fold cross-validation. In addition to the basic train/test split from Part I, you have to improve your evaluation across the different classes using *k-fold cross-validation*.[4] Perform a 10-fold cross-validation evaluation (with random shuffling) on your AI and add the results to your report. Make use of *scikit-learn*/skorch for splitting your datasets (i.e., do *not* use a manual, static split).[5]

Report. Update your report, correcting any issues from the first version and adding a new chapter for the bias detection & elimination, as well as new evaluation sections for the cross-validation.

Bias: Describe which of the above attributes you are analyzing and how. Provide evaluation metrics (P/R/F and confusion tables) for the different sub-classes (e.g., different age groups). Describe how the bias was introduced in your AI, how you addressed and re-trained it and compare the performance between the model from Part I and your new model.
Length: ca. 2 pages (incl. tables)

K-fold cross-validation: Add the results of your 10-fold cross-validation for both the original model (from Part I of the project) and your re-trained model (after the bias analysis and re-training). Compare the results with your previous, standard evaluation (fixed training/test split). Include the results of each individual fold, as well as the aggregate statistics, for precision, recall, F1, and accuracy. Provide any insights regarding your system's performance in an analysis section (i.e., explain differences between k-fold evaluation and your original train/test split evaluation).
Length: ca. 1 page

Deliverables. The structure of the deliverables is the same as for Part I (with your updated resources and report, of course). Submit the complete project, not just the changes you did for Part II.

Submission. Submit the updated version of your project through Moodle by the due date (late submission will incur a penalty, see Moodle for details).

Demo. We will schedule online demos sessions for your project using Zoom.

---

[1] E.g., *"Amazon scraps secret AI recruiting tool that showed bias against women"*, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08

[2] If your group has only two members, you have to analyze only one category

[3] See https://www.media.mit.edu/projects/gender-shades/overview/ as an example for systematic AI product testing.

[4] See https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation-iterators

[5] See https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html