Dear Sir or Madam,

Much appreciate for your invaluable suggestions. My answers are as follows,

Comment 1: First, I'm not sure that this problem is really a problem. I do regular searches for faculty home pages all the time and rarely have any problem finding them that can't be solved by appending the name of the university to the name of the faculty member. Why do we need a special-purpose people search engine (which the authors are apparently building) for this task? The ability to extract and analyze demographic information is more interesting, and the authors may have been better off writing a paper focused more on that aspect, but this isn't it. Right now, I can't see that this problem is of any substantial practical or theoretical interest.

Response: The general search engine can be queried with a faculty name to retrieve relevant information. However, it is hard to answer a simple question like how many computer science researchers in Yale University or the United States. As the reviewer pointed out that it is valuable to analyze the demographic information of the faculty members. Such kind of questions are even sort of sensitive to national security strategies. For instance, how many researchers are focusing on the quantum computing or nanotechnology, who are they, and what are estimated numbers in the near future based on the historical increasing? Professor++, a people-oriented search engine proposed in this study, is able to provide such knowledge from extracted faculty member databases. Such knowledge can be further mingled with an advanced academic question & answering system. In addition, we have developed an app (Face++) to help academic conference participants to know the session speakers. Face++ identifies a professor and provides her/his relevant information to users by taking a photo and searching our faculty member databases via a face recognition technique.

In addition, recognizing faculty homepages is a good example of the multimodal classification problem. The proposed multimodal generative and fusion framework is able to be generalizable to many other multimodal learning problems with class-imbalanced data and interactive feature modes. A good example is the prediction of media-aware stock movements, in which the market information space consists of several interactive modes, including transaction data, news articles, and investors' mood [1]. In bear markets, most stocks have downward pressure, that is, most samples are negatives which lead to a serious data imbalanced challenge.

[1] Qing Li, Yuanzhu Chen, Li Ling Jiang, Ping Li, and Hsinchun Chen. A tensor-based information framework for predicting the stock market. ACM Transactions on Information Systems (TOIS), 34(2):11, 2016.

Comment 2: Second, I can't see anything obviously novel in the classification methods, which might be overkill for this problem. We have some baselines in Table 2, but with not enough information to make them believable. To be convinced, I would need to see some carefully established baselines, with full details about the software

packages used, training, tuning, etc.

Response: Much appreciate for this invaluable suggestion. Due to the limited pages, we did not elaborate the details of model training and tuning in the submitted manuscript. However, we would like to add more details in the revised version. In addition, we have released a full version of our work which has not been submitted to anywhere else for publication. This full version can be accessed at https://github.com/mrspider520/gated_fusion_network/blob/master/paper/paper.pdf. It provides extra experiments about model tuning and model comparisons with different data sets in terms of scale and imbalance degree. The proposed framework shows the advantage to deal with larger and more imbalanced data with multimodal interactions. We also released the data and source code with this publication, in which the details about model training and tuning were presented. Besides, We have disclosed the source code along with the experimental dataset via Github website which can be accessed at https://github.com/mrspider520/gated_fusion_network.git

Comment 3: Third, the test collection consists of 31,645 web pages from "several official" university websites. The authors crawled the collection themselves. How? Why these pages? Clearly these can't be exhaustive crawls of these university domains. How do we know that all faculty homepages at these universities are captured in this crawl? Clearly some potential pages are filtered by the crawler, which is then doing some of the classification work. What are the details? At many places faculty may have two or more home pages (e.g., an official department one, a more personal research-focused one, another because they are director of some institute, or something like that). There are at least three pages at my university that could be considered my "home page". Do the authors have any criteria for selecting the correct one?

Response: Crawling efficiency & classification performance are two different tasks. In this research, we focus on the multimodal data classification which distinguishes faculty homepages from the crawled webpages. BTW, we applied the patented semi-focused crawler to construct the webpage corpus for evaluation. This patented semi-focused crawler is able to search university websites with pre-defined path rules (e.g. university webpage => college webpage => school/department webpage => faculty homepage) to download faculty homepages which greatly improves the crawling efficiency, and reduces the crawled webpages. The crawling efficiency is evaluated in another work.

In addition, we do not unify the faculty members with several homepages as our task is to differentiate between faculty homepages and non-faculty webpages. To some extent, users would prefer to have a full understanding of a certain faculty member by returning all of his/her homepages. However, it would be an interesting topic to identify multi-homepage faculty members in our future work.