

## Abstract

1. Reddit project (Meeting with Sina and Kevin - find a way to treat continuous attributes)
2. Display Advertising project (Meeting with Pegah Co - Asking my expected contribution)

## Description

### 1. Reddit project (Meeting with Sina and Kevin - find a way to treat continuous attributes)

We had a meeting with Sina and Kevin. We discussed major tasks of our research and presented the work done so far to them. We decided to work on gathering data not from network alignment, but just using RegEx and NLP methods. Currently we are working on gathering data (and its really difficult for attributes like age and location)

Also i talked with Armin about this problem and found a good way to deal with continuous attributes. We can quantize the attribute value to discrete bins and then use some smoothing kernel like Gaussian (RBF) kernel to simulate similarities between close bins and dissimilarities between far bins. I hope i can implement it to be used for age attribute. But unfortunately we haven't gather enough data for this attribute and we should wait until we get sufficient data.

### 2. Display Advertising project (Meeting with Pegah Co - Asking my expected contribution)

Finally, we had our meetings started with Abbas and Ali and we discussed about this project. They told me the main goal of this research should be solving the cold start problem for CTR / CVR prediction task. Also they told me i must read more papers. I have to find the actual metric and tree of important papers (But i think it should be a really time consuming task for me and it will take some time.) Also they told me i must use OutBrain RTB data as the main dataset. I downloaded it (36GB) but i couldn't unzip it (around 100GB) and it would be a really big problem.

## Next Week

- Reddit project (Meeting with Sina - Gather data for location, Age and Education attributes)
- Display Advertising project (Meeting with Pegah Co, talk more about metrics and important papers - Read more papers - Search ADKDD papers for relevant work - Look at OutBrain data characteristics)
- SP course TA work (Define homeworks and quizzes)

## References