- **Interesting papers:**
  - [2]A Habit Mining Approach for Discovering Similar Mobile Users (2012): They used contextual mobile data to try to regroup users by similar behaviors
    - Similar dataset than ours
    - Tried to capture user's behaviors and that's what we plan to do
  - [7]Exploiting Enriched Contextual Information for Mobile App Classification(2013): They tried to classify mobile apps by combining user mobile data, with some web data about the apps
    - a part of the project is dealing with a similar dataset than our
    - Algorithm that they found heavily outperforms the existing ones so may be can find a good idea in their algorithm that could be applied to our problem
  - Mining Personal Context-Aware Preferences for Mobile Users(2012): They are discovering users behaviours or preferences by combining the users records
    - same dataset
    - good results, maybe useful to see what they have done
  - Social Media Mining and Social Network Analysis: Emerging Research Guandong Xu (University of Technology, Sydney) and Lin Li (Wuhan University of Technology): A book released in 2013 that is cited in the three papers above. It contains a whole chapter explaining how to deal with the emergence of the new mobile data and how to extract similar user behaviors from this mobile data.
    - Need to purchase it, no free version found
  - [1]An Effective Approach for Mining Mobile User Habits (2010): tackles a problem very close ours.
  - [6]Context-Aware Role Mining for Mobile Service Recommendation(2012): Tries to regroup users by same context behavior to be able to recommend services to them. Example is if you know that user listens to pop music in the mornings, recommend a known pop music app to him.
  - Mining frequent patterns without candidate generation (2000): search for this one, seems to be interesting as the starting art (cited in many papers)
  - [3] Discovering significant places from mobile phones(2009)
  - [4] context-aware query classification(2009)
  - [5]Linearly constrained Bayesian matrix factorization(2009)
  - [8] V. Etter, M. Kafsi, E. Kazemi, M. Grossglauser, P. Thiran Where to go from here? Mobility prediction from instantaneous information (2013)
    - Using Nokia dataset (similar than ours with more users), developed a new model that abled to predict the next place to which to user will be
    - Outperformed the current existing models and won the Nokia challenge

- **Pertinent Papers to our work:**
    - **Problems addressed**

Cao et al. [1] addressed in 2010 the following problem: having a dataset of Nokia smartphone traces, they tried to discover the context that causes a user to have a certain interaction with the phone. An interaction could be "listening to music", "reading news", "having a message session". A context is a set composed by the features that are not phone interactions. For example a Context is C= {"Is holiday=Yes", "Day period=morning", "Phone profile= silent", "speed =High"}. They were for example interested in learning that the context C usually implies the interaction "reading news".

In extension to this work, Ma et al. [2] tried to detect similarities between users based on their habits (2012). Based on the same dataset and taking the result of [1] as input, they tried to cluster the users by behavior similarities. This work could be used for context-aware recommendation systems. Context aware recommendation systems take into account the context of the user to give him the right recommendation at the right time.

More work have been done with this kind of dataset as [6]and [8], however we limit our description to the researches described below because it is the most relevant to our work.

- **Dataset Cleaning**

In [1], the results presented show that the dataset is categorized. The interactions are categorized into "listening to music", "listening to visible radio", ect… Different categories of time ranges are selected and superposed in the same context as for example {"Is not holiday", "Day period = Noon", "Time range = PM3:00-4:00", "Day name"=Thursday}. The location of the user is decided by the cell to which he is connected to. However, no information about the way how this categorization is done is given. Nevertheless, it gives us insight about how our data should be categorized in order to be able to see meaningful results.

In [2], one of the main challenges is to overcome the sparsity of the data. In fact, the individual users' behaviors are very different from one user to the other even if they represent the same semantic. Thud, a big challenge is to be able to categorize and discretize the data so that similar behaviors between users could be caught. First of all, to reduce the sparsity of the locations, they decide to label each user's locations into one of the three categories: "Home", "Work Place" and "Others". For that they use an approach proposed by Yang [3], which discovers automatically those three clusters for each user. Second, to reduce the sparsity of the user-phone interactions they use a set of 13 predefined interactions defined by Nokia Ovi store (www.ovi.com). To map the initial interactions into those categories, they use the work introduced in [4] which only needs a small seed of labeled interactions to be able to automatically classify the interactions into the target categories. Related to our work, we especially keep in mind [3] and [4] which can be useful if we need to reduce the sparsity of our data.

- **Analysis Methods used**

In [1], they use association rules to learn the contextual information that leads to a phone interaction. The idea is to look through all the contextual features, build contexts and count how many times a context appears with a certain type of phone interaction. As checking all the combinations of the contextual features exponentially explode, thresholds min_support and min_confidence are used to only go through the promising contexts.
The main critic we have on this work is that by dividing initially the features into interaction and contextual features, we can learn only context that lead to an interaction with a phone. Moreover, interaction features at some time can be contextual at another time. In fact, a present user interaction can be caused by a past interaction. For example, when the user is in the train, he always opens his mail (interaction 1) then reads the news (interaction 2). In this case the context {"In the train = Yes", "mail"} leads to the interaction "reading news". In this work, the interaction features are never considered as potential contextual features and a property as the one described above can never be detected.

In [2], they use a constrained based matrix factorization model described in [5] to extract super-behaviors (which are clusters of behaviors). This method is a matrix factorization method that allows using some prior constrains and that gives as an output a distribution of factor matrixes. We keep this method in mind and we may want to understand it better when we will begin the analysis part. Especially, we want to know how this method differs from a usual probabilistic model and what are its weaknesses and strengths with respect to such a model.