Research Project Proposal: Multiple Source Feedback Clustering

Francesco Fulco Gonzales, francesco.gonzales@mail.polimi.it

1. Introduction to the problem

The goal of this research is to define a framework that allows to perform an online learning task on a set of objects with sparse interaction data, by compensating the data scarcity with a clustering approach. The objects of clustering can be described by unstructured multi-source data such as text and images and the clustering will have to be determined by the user interaction patterns with such objects.

In order to solve the above problem, a new feedback clustering framework is proposed. The main idea is that objects in the same cluster behave in the same way w.r.t. the online learning task, and therefore the interaction data of few objects, are descriptive of other objects in the same cluster, on account of such similarity. Moreover, the clustering distance metric can be learned through the interaction data collected in the online task, effectively boosting the task performance by changing the predicted clustering. The task can be therefore effectively performed on objects that have few or no interactions. Since the task is online, a way to update clusters as new data is collected should be devised, thus creating a feedback on the learned clusters, and similarity between objects.

To cluster the objects, and therefore learn a similarity between objects, a three-stage architecture is proposed:

1. Initially, objects are clustered using multiple and possibly unstructured information sources, which is used for initialization of the clusters in a task-independent fashion.
2. The baseline clustering of objects is then modified by incorporating the user interaction data on the specific task to be solved.
3. The clusters are continuously updated in an online reinforcement learning setting [13] as new interactions with objects flow in.

In the setting defined above, the chosen task can be successfully performed even on low or zero interaction objects by relying on close interaction-rich objects. The underlying intuition behind the proposed framework is that the representations extracted from the initial task-independent clustering is general enough to provide a good starting point to the subsequent task.

Case study This framework proves to be particularly useful for applications that lack sufficient interaction data on many data points, making difficult to apply current conventional methods. Considering this insight, the task chosen to test the framework outlined above is pricing in an e-commerce setting, where the objects are represented by products and the interactions are prices at which each item was proposed to a user, along with a boolean variable describing whether the object was then bought by the user. The goal of the task is to select the optimal price for each product in order to maximize profits. The proposed framework is particularly relevant in this setting as the interaction data are highly concentrated on few products and very scarce for the overwhelming majority of products. Therefore, the available information consists in product pictures and technical sheets, that are used to perform the initial clustering, and the sparse past interaction data for the task-specific clustering refinement. The final algorithm will finally be deployed to test its online learning capabilities to further update the clusters to incorporate the newly generated data of new purchases.

Assumptions The main assumptions the research project relies on are:

1. Object images contain a significant amount of information that can be leveraged for clustering.
2. The similarity between objects carries over the similarity for a generic task, i.e. initial clustering learned from the object characteristic provides a notable starting point for the subsequent task-specific clustering.
3. The proposed framework is general enough to provide a feasible solution to many tasks.

Contributions As mentioned above, the scope of the solution aims to be as general as possible. An algorithmic framework to incorporate an online feedback as supervisory signal to clustering would be a significant contribution to the scientific community. Moreover, the commercial applications of such a result would be even more impactful. In fact, solving the proposed e-commerce case study alone would be a great feat and greatly increase the profitability of many businesses, given the ever increasing magnitude of online spending.

2. Main related works

Very few studies have been conducted with the goal of solving the problem presented above. However, the proposed solution puts together two main fields, that can be studied separately and integrated: Deep Clustering and Contextual Bandit [7], a generalization of Multi-Armed Bandit, a well-known Reinforcement Learning problem setting.

Deep Clustering (DC) is a very active area of research that investigates methods to perform clustering using deep learning techniques, which have been shown to yield state-of-the-art performance to unsupervised learning tasks, thanks to their ability to learn complex non-linear representations of data [4], which are subsequently much easier to cluster. There are a wide variety of DC methods, that make use of different deep learning architectures, ranging from standard Multi-Layer Perceptrons to Convolutional Neural Networks, Autoencoders, Generative Adversarial Networks and more. The clustering algorithms are also just as varied, and employ, among the others, partition-based methods like k-means or hierarchical methods such as agglomerative clustering (more on clustering in [14]). Several methods have been proposed to combine the main network task of learning the latent representation and the clustering task, which can be optimized in a joint loss, or performed separately. For a thorough survey of deep clustering methods please refer to [9, 2].

Multi Armed Bandit (MAB) [12] is a classic reinforcement learning problem that exemplifies the exploration–exploitation tradeoff dilemma [1]. In the MAB setting the agent is faced repeatedly with a choice among multiple actions. After each choice the agent receives a scalar reward sampled from a probability distribution that depends on the selected action. The objective of the agent is to maximize the expected total reward over some finite time period. Contextual Bandit [7] is a generalization of MAB in which, at each round, the agent also takes as input a feature vector, called the context vector, that is used together with the past observations to choose the arm to play. Over time, the agent will gather enough data about how context vectors and rewards relate to each other and will be able to predict the next best arm to play by using the feature vectors.

In the pricing setting, which is a common use case for MAB [10, 11, 5], a product cluster represents the context vector given as input to the learner, which is faced with the choice of the price for the given cluster. The reward will be given according to the user response to the chosen price for the products belonging to that cluster, e.g. proportional to the margin on that category of products. This reward will serve as supervisory signal to adjust both the chosen price (or any other generic action prescripted by the chosen task) and the clustering of objects, where the former objective is the explicit goal of the MAB and the latter can be achieved by changing the latent representation of input data by having the reward acting as feedback control on the neural network [8].

Recent works investigate performing clustering using MAB for recommendation [3] and pricing [6] tasks. However, to the best of our knowledge, no work addresses a unified framework to perform an online reinforcement learning task with scarce data that leverages and concurrently updates the input clusters with the supervisory feedback provided by the online learning task.

3. Research plan

The goal of the research is to develop a framework that performs clustering based on interaction data on the objects in an online feedback fashion. The contribution that the proposed project intends to provide is primarily algorithmic and experimental. Related works will be leveraged to implement a learning algorithm that will be tested and validated from both a formal and experimental point of view. The efficacy of the proposed solution will be tested on the practical case study of pricing, that perfectly fits the problem definition.

As shown in the GANTT diagram of Figure 1, the task will be composed of three macro tasks, namely:

Preliminary research will lie a foundational understanding of the field through the analysis of the state of the art, for both deep clustering and contextual bandit, which allow a clearer definition of the research question.

Clustering is the phase in which the state-of-the-art deep clustering algorithms will be implemented and evaluated, in order to have a baseline clustering performance on both public and private datasets. The clustering algorithms will be evaluated using the standard metrics such as *unsupervised clustering accuracy (ACC)* and *Normalized Mutual Information (NMI)*.

Feedback Reinforcement Learning will finally be designed, developed and integrated with the previously implemented clustering algorithm. This final phase is the major contribution of the research project as it will implement the feedback system used to update the clustering and perform the actual pricing task. The system will finally be deployed in an online learning environment to evaluate its performance.

The project will end with the writing of a paper that will be submitted to one of the main conferences in the field, and will be extended into a thesis.

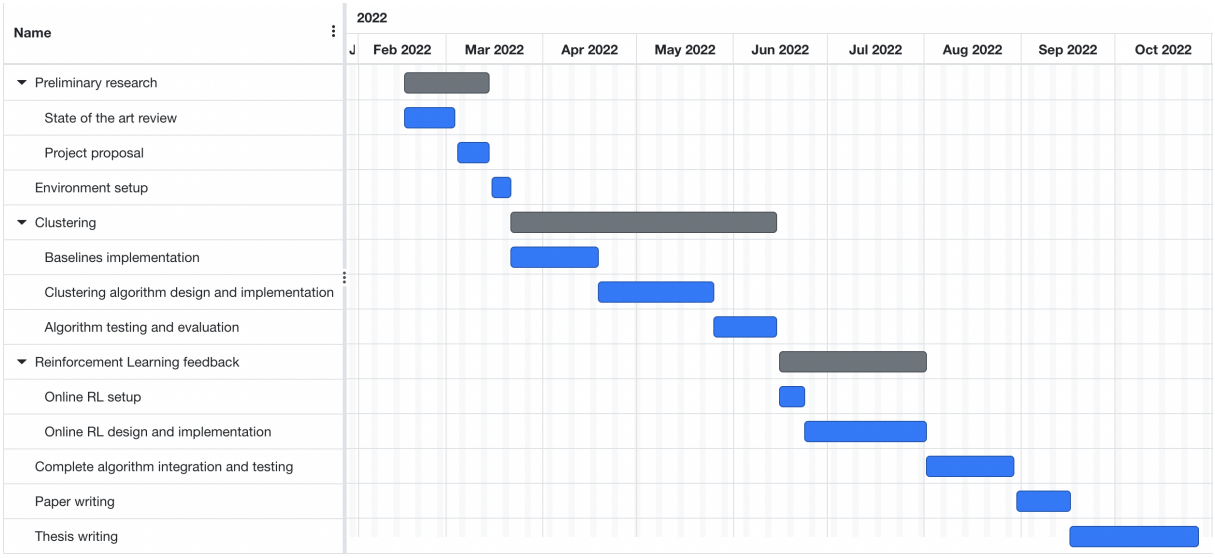

Figure 1: GANTT diagram

References

[1] Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems 24* (2011).

[2] Aljalbout, E., Golkov, V., Siddiqui, Y., Strobel, M., and Cremers, D. Clustering with deep learning: Taxonomy and new methods. *arXiv preprint arXiv:1801.07648* (2018).

[3] Ban, Y., and He, J. Local clustering in contextual multi-armed bandits. In *Proceedings of the Web Conference 2021* (2021), pp. 2335–2346.

[4] Bengio, Y., Courville, A., and Vincent, P. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence 35*, 8 (2013), 1798–1828.

[5] Bouneffouf, D., and Rish, I. A survey on practical applications of multi-armed and contextual bandits. *arXiv preprint arXiv:1904.10040* (2019).

[6] Genasti, G. A multi-armed bandit approach to dynamic pricing.

[7] Lu, T., Pál, D., and Pál, M. Contextual multi-armed bandits. In *Proceedings of the Thirteenth international conference on Artificial Intelligence and Statistics* (2010), JMLR Workshop and Conference Proceedings, pp. 485–492.

[8] Meulemans, A., Tristany Farinha, M., Garcia Ordonez, J., Vilimelis Aceituno, P., Sacramento, J., and Grewe, B. F. Credit assignment in neural networks through deep feedback control. *Advances in Neural Information Processing Systems 34* (2021).

[9] Min, E., Guo, X., Liu, Q., Zhang, G., Cui, J., and Long, J. A survey of clustering with deep learning: From the perspective of network architecture. *IEEE Access 6* (2018), 39501–39514.

[10] Misra, K., Schwartz, E. M., and Abernethy, J. Dynamic online pricing with incomplete information using multiarmed bandit experiments. *Marketing Science 38*, 2 (2019), 226–252.

[11] Mueller, J. W., Syrgkanis, V., and Taddy, M. Low-rank bandit methods for high-dimensional dynamic pricing. *Advances in Neural Information Processing Systems 32* (2019).

[12] Slivkins, A. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272* (2019).

[13] Sutton, R. S., and Barto, A. G. *Reinforcement learning: An introduction*. MIT press, 2018.

[14] Xu, D., and Tian, Y. A comprehensive survey of clustering algorithms. *Annals of Data Science 2*, 2 (2015), 165–193.