- *MT-DNN*[19] is an open-source natural language understanding toolkit by Microsoft to train customized deep learning models.

- *Yahoo Multi-Task Learning for Web Ranking* is a multi-task learning framework developed by Yahoo! Labs to rank in web search.

- *VIRTUAL* [12] is an algorithm for federated multi-task learning with non-convex models. The server and devices are treated as a star-shaped bayesian network, and model learning is performed on the network using approximated variational inference.

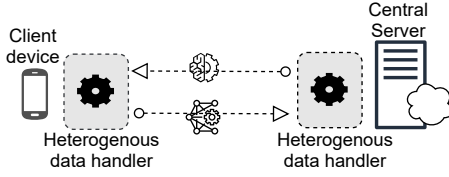### 3.3.2. Pattern 9: Heterogeneous Data Handler

Figure 12: Heterogeneous Data Handler.

Summary: Heterogeneous data handler solves the non-IID and skewed data distribution issues through data volume and data class addition (e.g., data augmentation or generative adversarial network) while maintaining the local data privacy. The pattern is illustrated in Fig. 12, where the heterogeneous data handler operates at both ends of the system.

Context: Client devices possess heterogeneous data characteristics due to the highly personalized data generation pattern. Furthermore, the raw local data cannot be shared so the data balancing task becomes extremely challenging.

Problem: The imbalanced and skewed data distribution of client devices produces local models that are not generalised to the entire client network. The aggregation of these local models reduces global model accuracy.

Forces: The problem requires the following forces to be balanced:

- *Data efficiency.* It is challenging to articulate the suitable data volume and classes to be augmented to solve data heterogeneity on local client devices.

- *Data accessibility.* The heterogeneous data issue that exists within the client device can be solved by collecting all the data under a centralized location. However, this violates the data privacy of client devices.

Solution: A heterogeneous data handler balances the data distribution and solves the data heterogeneity issue in the client devices through data augmentation and federated distillation. Data augmentation solves data heterogeneity by generating augmented data locally until the data volume is the same across all client devices. Furthermore, the classes

---

¹⁹https://github.com/microsoft/MT-DNN

in the datasets are also populated equally across all client devices. Federated distillation enables the client devices to obtain knowledge from other devices periodically without directly accessing the data of other client devices. Other methods includes taking the quantified data heterogeneity weightage (e.g, Pearson's correlation, centroid averaging-distance, etc.) into account for model aggregation.

Consequences:

Benefits:

- *Model quality.* By solving the non-IID issue of local datasets, the performance and generality of the global model are improved.

Drawbacks:

- *Computation cost.* It is computationally costly to deal with data heterogeneity together with the local model training.

Related patterns: *Client Registry, Client Selector, Client Cluster*

Known uses:

- *Astreae* [13] is a self-balancing federated learning framework that alleviates the imbalances by performing global data distribution-based data augmentation.

- Federated Augmentation (*FAug*) [19] is a data augmentation scheme that utilises a generative adversarial network (GAN) which is collectively trained under the trade-off between privacy leakage and communication overhead.

- Federated Distillation (*FD*) [3] is a method that adopted knowledge distillation approaches to tackle the non-IID issue by obtaining the knowledge from other devices during the distributed training process, without accessing the raw data.
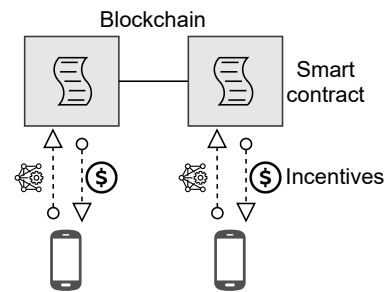
3.3.3. Pattern 10: Incentive Registry

Figure 13: Incentive Registry.

Summary: An incentive registry maintains the list of participating clients and their rewards that correspond to clients' contributions (e.g., data volume, model performance, computation resources, etc.) to motivate clients' participation. Fig. 13 illustrates a blockchain & smart contract-based incentive mechanism.