**Описание предполагаемого способа решения**

Кластеризация распределённых данных представляет некоторые сложности применительно к нейронным сетям, так как не позволяет произвести обучение сети до тех пор, пока все данные не будут собраны на одном центральном сайте.

На основе рассмотренных аналогов было решено сосредоточиться на использовании алгоритма Растущего Нейронного Газа (Growing Neural Gas) относящегося к классу Сетей, представляющих Топологию (Topology Representing Network), являясь потомком Самоорганизующихся карт имеет преимущество в возможности работы над динамически меняющимся набором данных. GNG – использует модель «победитель получает всё».

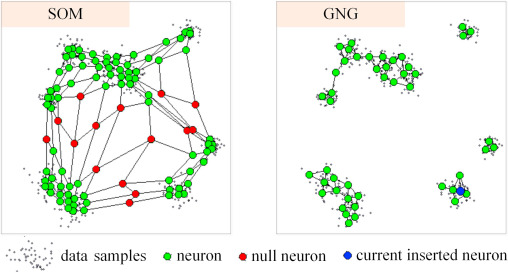


Рис. 1. Сравнение работы SOM и GNG

Ещё одним важным условием является необходимость обеспечить общий формат данных при переходе от локальных сайтов к глобальным, это также позволит справиться с ограничениями в виде индивидуальной и корпоративной конфиденциальности, можно будет передавать только матрицу разбиения локальных данных или использовать оценку плотности локальных данных.

**Инструменты разработки**

Для разработки алгоритма, который будет обеспечивать данное взаимодействие, будет использоваться библиотека машинного обучения scikit-learn для языка программирования Python, которая предоставляет:

* Простые и эффективные инструменты для интеллектуального анализа данных
* Доступная для всех желающих с возможностью многократного использования в различных контекстах
* Построена NumPy, SciPy, и matplotlib
* Открытый исходный код, коммерчески применим - лицензия BSD

Для обучения нейронной сети будут использоваться наборы данных пример: <https://www.kaggle.com/datasets> и <https://archive.ics.uci.edu/ml/datasets.html> , первоначально проверка работы алгоритма будет проведена над данными которые расположены локально, потом будет произведено над распределёнными данными с применением хранилищ данных или Network File System.

**Модель данных**

Соответственно входными данными алгоритма будут являться числовые значения, возможно определённые самостоятельно из которых будет определяться вес нейронов в сети и на этой основе определяться победитель для обучения.

Для обеспечения согласованности будет реализован механизм приведения данных к общему формату он будет реализован на центральном сайте алгоритма, на котором и будет проводиться кластеризация.