Определение типов сдвигов между клиентами федеративного обучения в условиях горизонтального разделения данных

Д. А. Фомичев[1], И. И. Холод[1], А. А. Кочешков[2]

[1]*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)*

[2]*ООО «Смартилайзер Рус»*

dafomichev@etu.ru, iiholod@etu.ru, kocheshkovGIA@yandex.ru

***Аннотация*. В данной работе рассматривается задача определения типов распределений между клиентами федеративного обучения в условиях горизонтального разделения данных. Актуальность исследования обусловлена тем, что неоднородность локальных выборок клиентов может существенно влиять на качество обучения и обобщающую способность глобальной модели. В работе анализируются основные типы распределения данных между клиентами, включая сдвиг признаков, сдвиг меток и сдвиг концепции, а также предлагается метод к их выявлению на основе сравнения распределений локальных данных без передачи самих данных между клиентами.**

Ключевые слова: федеративное обучение, сдвиги данных, горизонтальное разделение данных

I. Введение

В парадигме федеративного обучения статистическая неоднородность локальных данных является основным фактором возникновения эффекта расхождения локальных моделей, что приводит к нестабильности агрегации и деградации обобщающей способности глобальной модели. В настоящее время разработан ряд решений, способный бороться с неоднородностью [1–2], однако в условиях федеративного обучения одной из важных задач является установление различного распределения в данных [6], при этом сохранив их конфиденциальность между клиентами, а также не увеличить объем сетевого трафика.

Для количественной оценки этой неоднородности часто применяются метрические меры, в частности, в исследованиях применяется расстояние Вассерштейна. Так, алгоритм FedWaD [3] позволяет оценивать глобальное различие распределений в федеративном режиме без раскрытия сырых данных. Суть метода заключается в итеративном приближении расстояния с помощью промежуточных распределений. Сервер отправляет текущее приближение ξ, клиенты локально вычисляют интерполирующие меры между своим распределением и ξ, после чего отправляют их обратно серверу. Сервер на основе этих промежуточных мер уточняет свое приближение. Однако, обеспечивая точную численную оценку близости, подобные метрические подходы не позволяют определить конкретный тип распределения данных между клиентами.

Работа выполнена при поддержке гранта Российского научного фонда, №25-11-20020 (https://rscf.ru/project/25-11-20020) и Санкт-Петербургского научного фонда

Предлагаемый в настоящей работе параметрический метод дополняет существующий инструментарий анализа, позволяя не только фиксировать наличие неоднородности, но и напрямую выявлять конкретные типы распределений: сдвиг признаков, сдвиг меток и сдвиг концепции. Использование агрегированных статистик обеспечивает минимальную коммуникационную нагрузку на сеть и предоставляет высокую степень интерпретируемости, необходимую для адаптации стратегий обучения под конкретный вид распределения данных.

Рассмотрим задачу анализа распределений между двумя клиентами в федеративном обучении, которую можно интерпретировать как задачу анализа неоднородных данных. Пусть каждый клиент $i \in \{A, B\}$ предоставляет выборку с признаками X^{i} и метками Y^{i}. Согласно классификации, предложенной в работе [4] выделяют три основных типа распределения между наборами данных, которые характеризуются тем, какие компоненты совместного распределения $P(X^{\mathrm{i}}, Y^{\mathrm{i}})$ изменяются между клиентами, где маргинальное распределение – это распределение одной переменной без учета другой, например $P(X)$ для признаков или $P(Y)$ для меток, а условное распределение – распределение одной переменной при условии, что вторая принимает определенные значения:

- сдвиг признаков — отличие маргинального распределения признаков $P(X^{\mathrm{A}}) \neq P(X^{\mathrm{B}})$ при сохранении условного распределения $P(Y^{\mathrm{A}} \mid X^{\mathrm{A}}) = P(Y^{\mathrm{B}} \mid X^{\mathrm{B}})$;
- сдвиг меток — отличие маргинального распределения меток $P(Y^{\mathrm{A}}) \neq P(Y^{\mathrm{B}})$ при сохранении условного распределения признаков $P(X^{\mathrm{A}} \mid Y^{\mathrm{A}}) = P(X^{\mathrm{B}} \mid Y^{\mathrm{B}})$;
- сдвиг концепции — отличие условного распределения $P(Y^{\mathrm{A}} \mid X^{\mathrm{A}}) \neq P(Y^{\mathrm{B}} \mid X^{\mathrm{B}})$, то есть самой зависимости метки от признаков.

Анализ основан на двух эквивалентных разложениях совместного распределения:

$$P(X^{\mathrm{i}}, Y^{\mathrm{i}}) = P(Y^{\mathrm{i}} \mid X^{\mathrm{i}})P(X^{\mathrm{i}}) = P(X^{\mathrm{i}} \mid Y^{\mathrm{i}})P(Y^{\mathrm{i}}).$$

Поскольку совместное распределение полностью определяется через условное и маргинальное распределения, изменение одной из этих компонент позволяет классифицировать тип сдвига.

II. Описание типов распределения данных

Согласно исследованию [4] выделяют три основных типа распределения данных, которые можно определить как: сдвиг признаков (ковариантный сдвиг), сдвиг меток (априорный сдвиг) и сдвиг концепции.

A. Сдвиг признаков

Сдвиг признаков возникает между клиентами A и B, если маргинальные распределения признаков различаются:

$$P(X^{\mathrm{A}}) \neq P(X^{\mathrm{B}}),$$

при этом условное распределение метки остается неизменным:

$$P(Y^{\mathrm{A}} \mid X^{\mathrm{A}}) = P(Y^{\mathrm{B}} \mid X^{\mathrm{B}})$$

При ковариантном сдвиге главным признаком является изменение маргинального распределения признаков, тогда как условное распределение остается неизменным. Поэтому распределение меток $P(Y)$ не используется как определяющий критерий этого типа сдвига: оно может оставаться тем же или изменяться в зависимости от того, как меняется распределение признаков. Кроме того, в общем случае меняется и распределение признаков при фиксированной метке $P(X \mid Y)$, что следует из формулы Байеса [6]:

$$P(X^{\mathrm{i}} \mid Y^{\mathrm{i}}) = \frac{P(Y^{\mathrm{i}} \mid X^{\mathrm{i}})\, P(X^{\mathrm{i}})}{P(Y^{\mathrm{i}})}$$

B. Сдвиг меток

Сдвиг меток имеет место, если маргинальные распределения меток различаются:

$$P(Y^{\mathrm{A}}) \neq P(Y^{\mathrm{B}}),$$

при сохранении условного распределения признаков при фиксированной метке:

$$P(X^{\mathrm{A}} \mid Y^{\mathrm{A}}) = P(X^{\mathrm{B}} \mid Y^{\mathrm{B}}).$$

Из формулы полной вероятности $P(X^i) = \sum_y P(X^i \mid Y^i = y) P(Y^i = y)$ следует, что при неизменном $P(X^{\mathrm{i}} \mid Y^{\mathrm{i}})$ изменение $P(Y^{\mathrm{i}})$ в общем случае влечёт изменение маргинального распределения признаков $P(X^{\mathrm{i}})$.

C. Сдвиг концепции

Сдвиг концепции характеризуется различием условного распределения метки по признакам между клиентами:

$$P(Y^{\mathrm{A}} \mid X^{\mathrm{A}}) \neq P(Y^{\mathrm{B}} \mid X^{\mathrm{B}}).$$

В качестве частного, упрощающего случая часто рассматривают ситуацию, когда маргинальные распределения признаков и меток совпадают:

$$P(X^{\mathrm{A}}) = P(X^{\mathrm{B}}), P(Y^{\mathrm{A}}) = P(Y^{\mathrm{B}}),$$

однако это не является обязательным условием.

Сдвиг концепции означает изменение самой зависимости между признаками и метками, то есть функции принятия решения. Иными словами, даже если маргинальные распределения могут частично совпадать, условная связь $P(Y^{\mathrm{i}} \mid X^{\mathrm{i}})$изменяется, что отличает сдвиг концепции от сдвига признаков или сдвига меток. На практике этот тип сдвига часто сопровождается и другими изменениями распределений, поэтому он считается наиболее сложным для анализа и самым серьезным сдвигом при обучении.

III. Параметрический метод вычисления и сравнения распределений

Для вычисления и сравнения распределений между клиентами предлагается унифицированный метод на основе непрерывных распределений. Такой метод позволяет сравнивать распределения клиентов по агрегированным статистикам, не передавая исходные данные.

A. Сравнение распределений между клиентами

Сравнение выполняется не по исходным данным, а по оцененным параметрам распределений и по значениям вероятностей, восстановленных по формуле Байеса.

Анализ распределения признаков и последующая оценка между двумя клиентами A и B выполняется следующим образом: для каждого признака $\mathrm{X_r}$ рассчитывается разность их математических ожиданий и стандартных отклонений:

$$d_{\mu}(X_r) = \left|\mu_{X_r}^{(A)} - \mu_{X_r}^{(B)}\right|$$
$$d_{\sigma}(X_r) = \left|\sigma_{X_r}^{(A)} - \sigma_{X_r}^{(B)}\right|$$

Нормированные разности:

$$\widetilde{\mathrm{d_{\mu}}}(\mathrm{X_r}) = \frac{\mathrm{d_{\mu}(X_r)}}{\max\left(\mathrm{R_{\mu}^{A}}, \mathrm{R_{\mu}^{B}}\right)}; \widetilde{\mathrm{d_{\sigma}}}(\mathrm{X_r}) = \frac{\mathrm{d_{\sigma}(X_r)}}{\max\left(\sigma_{\mathrm{X_r}}^{(A)}, \sigma_{\mathrm{X_r}}^{(B)}\right)}$$

$\mathrm{R_{\mu}^{A,B}} = \mathrm{x_{max}^{A,B}} - \mathrm{x_{min}^{A,B}}$ – диапазон данных клиентов (минимальная и максимальная граница или закодированные категории).

Используем нормированную метрику, так как масштабы параметров признака разные.

Тогда нормированная метрика равна:

$$D_X = \sqrt{\widetilde{d_{\mu}^2} + \widetilde{d_{\sigma}^2}} \in [0, \sqrt{2}]$$

Пороговые значения определяются как квантили максимального расстояния:

- [0; 0.35) – незначительное (25%);
- [0.35; 0.7) – умеренное (50%);
- [0.7; 1.05) – значительное (75%);
- > 75% – критичное.

Судить о наличии сдвига признаков можно уже при умеренном различие, в частности условное распределение (сравнение которого представлено далее) будет стремиться к минимуму, если речь идет только о сдвиге признаков. Это уже дает понять, что данные между клиентами в некоторой степени различаются, что может отразиться на последующем ухудшении точности глобальной модели, из-за замедления глобальной сходимости [3]. Такой подход к нормализации согласуется с практикой сравнения моделей в нормированных пространствах [7–8].

Для сравнения распределения меток используется абсолютная разность априорных вероятностей (не нуждается в нормировке, так как вероятности уже имеют

универсальный масштаб и не могут выходить за пределы диапазона [0;1]):

$$d_Y(y_m) = \left|P^{(A)}(Y = y_m) - P^{(B)}(Y = y_m)\right|$$

Комплексная метрика представляет собой полное вариационное расстояние, и по свойству вероятностей находится в диапазоне значений [0,2]:

$$D_Y = \sum_{m=1}^{M} d_Y(y_m) = \sum_{m=1}^{M} \left|P^{(A)}(Y = y_m) - P^{(B)}(Y = y_m)\right|$$

Тогда пороговые значения для определения различия в распределениях меток:

- [0; 0.5) – незначительное (25%);
- [0.5; 1) – умеренное (50%);
- [1; 1.5) – значительное (75%);
- > 75% - критичное.

Судить о наличии сдвига меток можно при умеренном различии, разные доли классов могут в последствии вызвать несбалансированную агрегацию весов. При этом различие в условных распределениях $\mathrm{P(X \mid Y)}$ при чистом сдвиге меток будут минимальны.

Сравнение условных вероятностей $\mathrm{P(Y \mid X)}$ для последующего определения сдвига концепции между клиентами проводится по фиксированной сетке значений признака:

$$d_{Y|X_r(x_l)} = \sum_{m=1}^{M} \left|P^{(A)}(Y = y_m \mid X_r = x_l) - P^{(B)}(Y = y_m \mid X_r = x_l)\right|$$

где $x_1, \dots, x_L$ – равномерная сетка по диапазону значений признака X_r. Сетка значений формируется на сервере исходя из минимального и максимального значения признака. Большее количество точек в сетке значений обеспечивает лучшую оценку, но занимает больше времени.

Общая нормированная метрика (отражает среднюю величину различий) также лежит в диапазоне [0;2]:

$$D_{Y|X} = \frac{1}{L}\sum_{l=1}^{L} d_{Y|X}(X_l)$$

Уже при умеренном различии может нарушаться согласованность предсказаний глобальной модели [5], вызванная наличием сдвига концепции.

B. Аппроксимация через плотность распределения

Локальные выборки каждого клиента на сервере можно аппроксимировать через функцию нормального распределения [8–9], которая в общем виде выглядит следующим образом:

$$f_{X_r}^{(i)}(x) = \frac{1}{\sigma_{X_r}^{(i)}\sqrt{2\pi}} \exp\left(-\frac{\left(x - \mu_{X_r}^{(i)}\right)^2}{2\left(\sigma_{X_r}^{(i)}\right)^2}\right)$$

Это позволяет представить исходные выборки в виде компактного параметрического описания, сохраняющего основную информацию о центре и разбросе данных. В том числе, полученные оценённые плотности можно подставить в формулу Байеса, что позволит восстановить апостериорные вероятности $\mathrm{P(Y \mid X_r)}$ для каждого значения признака:

$$\mathrm{P^{(i)}(Y = y_m \mid X_r = x)} = \frac{\mathrm{A_m(x)}}{\sum_{j=1}^{M} \mathrm{A_j(x)}},$$

где

$$A_m(x) = f_{X_r|Y=y_m}^{(i)}(x) P^{(i)}(Y = y_m)$$

здесь $f_{X_r|Y=y_m}^{(i)}$ – оценённая плотность признака $\mathrm{X_r}$ при значении метки $\mathrm{y_m}$, $\mathrm{P^{(i)}(Y = y_m)}$ – априорная вероятность этого класса на клиенте i.

Таким образом, сравнение клиентов выполняется не по исходным наблюдениям, а по параметрам аппроксимированных плотностей и по восстановленным условным вероятностям, что позволяет ограничить количество передаваемых статистических параметров от клиентов на сервер.

C. Минимальный набор параметров для передачи

Исходя из описанного ранее метода для оценки распределений между клиентами, каждому из них достаточно передать следующий набор параметров:

Для каждого признака X_r независимо от типа:

- μ_{X_r}— математическое ожидание;
- σ_{Xr} — стандартное отклонение;
- минимальное и максимальное значение

Для каждого условного распределения $P(X_r \mid Y = y_m)$:

- $\mu_{X_r \mid y_m}$ — условное математическое ожидание;
- $\sigma_{X_r \mid y_m}$ — условное стандартное отклонение;

Для каждой метки:

- $P(Y = y_m)$ для каждой категории y_m.

Математическое ожидание и стандартное отклонение для числовых признаков вычисляется следующим образом:

$$\mu_{X_r}^{(i)} = \frac{1}{N_i}\sum_{n=1}^{N_i} x_{r,n}^{(i)}$$

$$\sigma_{X_r}^{(i)} = \sqrt{\frac{1}{N_i - 1}\sum_{n=1}^{N_i}\left(x_{r,n}^{(i)} - \mu_{X_r}^{(i)}\right)^2}$$

Для дискретных признаков с категориями $\{x_1, \dots, x_K\}$ необходимо предварительно провести кодировку признаков в числовые значения, затем использовать закодированные значения признака:

$$\mu_{X_r}^{(i)} = \sum_{k=1}^{K} k \cdot P^{(i)}(X_r = x_k),$$

$$\sigma_{X_r}^{(i)} = \sqrt{\sum_{k=1}^{K}\left(k - \mu_{X_r}^{(i)}\right)^2 \cdot P^{(i)}(X_r = x_k)},$$

где

$$P^{(i)}(X_r = x_k) = \frac{N^{(i)}(X_r = x_k)}{N_i}.$$

Для условного распределения $P(X_r \mid Y = y_m)$ параметры вычисляется аналогично, но для каждой метки $y_m \in \mathcal{Y} = \{y_1, \dots, y_M\}$, то есть математическое ожидание и стандартное отклонение признака при конкретном значении метки.

Для дискретных признаков условные параметры вычисляются через условные вероятности – это необходимо для вычисления условного математического ожидания и стандартного отклонения на клиенте:

$$P^{(i)}(X_r = x_k \mid Y = y_m) = \frac{N^{(i)}(X_r = x_k, Y = y_m)}{N_{y_m}^{(i)}}$$

D. Общая схема метода

Общую схему метода, позволяющего определить распределение данных между клиентами, а также идентифицировать характер сдвига при его наличии, можно описать следующим образом:

- Согласование кодировки: необходимо установить единую кодировку для дискретных признаков между клиентами, так как далее это необходимо для корректного получения статистических параметров. Без согласования основные параметры такие как математическое ожидание и стандартное отклонение будут несопоставимы.
- Локальное вычисление параметров: каждый клиент по своей выборке вычисляет параметры безусловных и условных распределений, необходимых в дальнейшем для их оценки на сервере и восстановления распределения с помощью аппроксимации для условных распределений.
- Передача параметров: на сервер передаются только агрегированные статистические параметры, а именно параметры распределений, а также объемы выборок. Это позволяет ограничить объем трафика, что особенно важно при федеративном обучении.
- Серверная аппроксимация: сервер восстанавливает непрерывные распределения на основе полученных параметров, используя нормальную аппроксимацию. Благодаря этому, можно восстановить апостериорную вероятность каждого клиента на сервере с помощью сетки значений признака
- Сравнение распределений: сравниваются параметры условных и безусловных распределений. Предложенные метрики нормированы для сопоставимости разнородных признаков с различными масштабами и единицами измерения, что обеспечивает универсальность метода.

Предложенный метод обеспечивает приватность данных клиентов, линейную вычислительную сложность, а также предоставляет интерпретируемую идентификацию распределений данных между клиентами. На основе нормированных метрик выполняется диагностика и обнаружение типа сдвигов, локализация конкретных признаков X_r с явно выраженной гетерогенностью.

IV. Заключение

Предложен параметрический метод сравнения распределений между клиентами федеративного обучения при горизонтальном разделении данных, использующий только агрегированные статистические характеристики. В общем случае, необходимы только математическое ожидание и стандартное отклонение для маргинальных и условных распределений, вероятность для каждого класса метки, а также минимальное и максимальное значение признаков. Данный метод позволяет определить наличие сдвига между локальными данными клиентов, не передавая сырые данные между ними или сервером. Кроме того, такая оценка позволяет локализовать источник сдвига как на уровне отдельного признака, так и с помощью комплексной оценки. Полученная информация о наличии сдвига распределений и «проблемных» признаках может быть напрямую использована для последующей предобработки локальных данных клиентов федеративного обучения, в частности для адаптивного выравнивания распределений.

Список литературы

[1] X. Ma, J. Zhu, Z. Lin, S. Chen, and Y. Qin, “A state-of-the-art survey on solving non-iid data in federated learning,” FGCS, vol. 135, pp. 244–258, 2022.

[2] W. Lu, J. Cheng, X. Li, and J. He, “A review of solving non-iid data in federated learning: Current status and future directions,” in Artificial Intelligence and Machine Learning, Singapore: Springer Nature Singapore, 2024, pp. 58–72.

[3] Rakotomamonjy A., Nadjahi K., Ralaivola L. Federated wasserstein distance //arXiv preprint arXiv:2310.01973. 2023.

[4] Moreno-Torres, J. G., Raeder, T., Alaiz-Rodríguez, R., Chawla, N. V., Herrera, F. (2012). A unifying view on dataset shift in classification. Pattern Recognition, 45(1), 521–530.

[5] Tan, X., Xie, T., Zheng, X., Yener, A., Lee, M., Payani, A., Latapie, H., & Zhang, X. (2026). Federated Learning Under Evolving Distribution Shifts. Entropy, 28(1), 101. URL: https://doi.org/10.3390/e28010101

[6] D.M. Jimenez G., D. Solans, M.A. Heikkilä, A. Vitaletti, N. Kourtellis, A. Anagnostopoulos, I. Chatzigiannakis, "Non-IID data in Federated Learning: A Survey with Taxonomy, Metrics, Methods, Frameworks and Future Directions", arXiv:2411.12377 [cs.LG], 2024.

[7] Benatti, A., Costa, L. da F. Normalization in Proportional Feature Spaces. arXiv:2409.11389 [cs.LG]. 2024 .URL: https://arxiv.org/abs/2409.11389

[8] Bishop, C. M. Pattern Recognition and Machine Learning. Springer, 2006.

[9] Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, New York, 2009. ISBN 978-0-387-84858-7.