

ПРОГНОЗИРОВАНИЕ КРАТКОСРОЧНОЙ ЛИКВИДНОСТИ КОММЕРЧЕСКОГО БАНКА С ИСПОЛЬЗОВАНИЕМ СЕТЕВЫХ МОДЕЛЕЙ

Егоркин А.А.

Российский Государственный Социальный Университет, Москва, Россия

2-5@bk.ru

Аннотация. В работе рассмотрены актуальные модели прогнозирования ликвидности коммерческого банка, а также подходы, базирующиеся на сетевых моделях. В качестве сетевых характеристик используются различные меры центральности узлов графа финансовых транзакций. С использованием алгоритмов машинного обучения показаны результаты применения предлагаемых моделей на реальных данных.

Ключевые слова: граф финансовых транзакций, прогнозирование ликвидности коммерческого банка, центральность узла графа.

Введение

Ликвидность банка – это способность банка обеспечить своевременное и полное исполнение своих обязательств. Управление ликвидностью банка направлено на предотвращение, в первую очередь, недостатка ликвидности. Недостаточная ликвидность может привести к неплатёжеспособности банка, а чрезмерная может снизить его прибыльность.

Среди существующих подходов к управлению и прогнозированию ликвидности можно выделить: нормативы ликвидности, GAP-анализ ликвидности и платежный календарь, сценарный анализ и стресс-тестирование риска ликвидности. Указанные выше подходы базируются, как правило, на существующей или прогнозной структуре баланса банка с применением к ней различных сценариев, в, например, стрессовых.

В настоящей работе предложен иной подход к прогнозированию краткосрочной ликвидности коммерческого банка с использованием сетевых моделей. В рамках данного подхода будет построен граф финансовых транзакций, представляющий собой сеть, описывающую денежные переводы в течение дня как внутри банка, так и приходящие в банк извне и уходящие из банка вовне. Исходя из динамики показателей центральности данного графа прогнозируется целевая переменная – объем ликвидных активов банка.

В качестве модели прогнозирования в настоящей работе рассматривается модель классификации, реализованная с помощью алгоритма машинного обучения – алгоритма случайного леса (RandomForestClassifier).

Все расчёты были проведены на реальных данных коммерческого банка, входящего в ТОП-40 российских банков по размеру активов, за период с 2022 по 2023 год. Поскольку большая часть информации может представлять собой коммерческую и банковскую тайну, все переменные и результаты были нормированы и обезличены.

1. Современные методы управления ликвидностью в коммерческих банках

Существующие методы анализа ликвидности коммерческого банка основываются в основной своей массе на текущей структуре баланса банка с использованием различных сценариев. Пример агрегированного баланса коммерческого банка с распределением требований и обязательств по срокам приведен в таблице 1.

Таблица 1. Пример агрегированного баланса банка и схема стресс-тестирования

Статья баланса	Доля в балансе	Оставшийся срок до погашения			
		1D	1D-1M	1-12M	>12M
Активы (размещенные средства)	100%				
1.1 Денежные средства и средства на к/с	4%	+	-	-	-
1.2 Краткосрочные межбанковские кредиты	5%	+	+	-	-
1.3 Долгосрочные межбанковские кредиты	1%	-	-	+	+
1.4 Ценные бумаги	18%	+	+	-	-
1.5 Кредитный портфель за вычетом резервов	63%	-	-	+	+
1.6 Прочие активы	10%	-	-	-	+
Обязательства (привлеченные средства и капитал)	100%				
Привлеченные средства	83%				
2.1 Средства банков привлеченные	1%	+	+	-	-
2.2 Средства ЮЛ	37%	+	+	+	-
2.3 Средства ФЛ	41%	-	+	+	+
2.4 Прочие обязательства	3%	+	-	-	-
Капитал	18%	-	-	-	+

К методам оценки ликвидности коммерческого банка относятся:

- Нормативы Банка России [1]:

$$H2 = \frac{\sum_{t \leq 1D} A_t}{\sum_{t \leq 1D} O_t - O_2^*} \geq 15\% \quad (1)$$

$$H3 = \frac{\sum_{t \leq 1M} A_t}{\sum_{t \leq 1M} O_t - O_3^*} \geq 50\% \quad (2)$$

$$H4 = \frac{\sum_{t > 12M} A_t}{\sum_{t > 12M} O_t + C + O_4^*} \leq 120\% \quad (3)$$

где:

A – активы;

O – обязательства;

C – капитал;

O_2^*, O_3^*, O_4^* – минимальные значения обязательств за последние 12 месяцев со сроками погашения: до востребования ($t \leq 1D$), до одного месяца ($t \leq 1M$), свыше года ($t > 12M$) – соответственно.

Нормативы $H2$ и $H3$ определяют минимальный размер ликвидных активов банка исходя из объема привлечённых краткосрочных обязательств, для того чтобы банк был в состоянии рассчитаться по ним. Норматив $H4$ ограничивает долгосрочное кредитование банка, чтобы долгосрочные кредиты (например, ипотека) выдавались за счет длинных привлечений.

- Анализ разрывов или GAP-анализ

$$GAP(T) = \sum_{t < T} A_t - \sum_{t < T} O_t \quad (4)$$

Величина $GAP(T)$ показывает кумулятивный дефицит/профицит ликвидности на различных временных интервалах (T). Банки могут накладывать внутренние ограничения на величину разрыва ликвидности на различных временных интервалах

- Стресс тестирование ликвидности [2].

В рамках данного подхода происходит одновременно удлинение срочности активов (например, за счет невозврата кредита) и сокращение срочности обязательств (например, за счет досрочного изъятия средств клиентами). С учетом используемого сценария банк оценивает риск ликвидности и возможность исполнить все свои обязательства перед клиентами.

Одним из вариантов регуляторного стресс-теста является норматив краткосрочной ликвидности (НКЛ) Банка России для системно значимых банков.

2. Целевая переменная и нулевая модель

В качестве целевой переменной для дальнейшего прогнозирования будем использовать величину ликвидных активов коммерческого банка. Ликвидные активы – это активы, которые могут в минимальные сроки быть превращены в деньги. К ним относятся следующие статьи баланса (таблица 1):

- Денежные средства и средства на корреспондентских счетах;
- Краткосрочные межбанковские кредиты (со сроком погашения до месяца);
- Ценные бумаги, которые могут быть быстро проданы, либо заложены по операциям РЕПО [3].

Целевая переменная нормирована на ее среднее значение за рассматриваемый период времени, т.е. среднее значение целевой переменной равно 1.



Рис. 1. Динамика нормированной целевой переменной за период 2022–2023 год

Будет решаться задача классификации, классифицируем целевую переменную следующим образом:

$$L_{class} = \begin{cases} class_1, L < 0.95 \\ class_2, 0.95 \leq L < 1.05 \\ class_3, L \geq 1.05 \end{cases} \quad (5)$$

Распределение по классам целевой переменной является сбалансированным (см. таблицу 2).

Таблица 2. Распределение по классам целевой переменной

L_{class}	Доля целевых переменных	Среднее	Медиана	σ
$class_1$	30.9%	0.89	0.91	0.06
$class_2$	34.5%	1.00	1.00	0.03
$class_3$	34.5%	1.14	1.12	0.07

Целевая переменная не имеет трендовой составляющей, ее движение происходит около среднего значения, что имеет экономическое объяснение. Объем ликвидных активов поддерживается не выше определенного уровня, поскольку они, как правило, имеют минимальную доходность. Таким образом, наиболее простым способом прогнозирования целевой переменной является использование ее текущего и исторических значений:

$$L_{class_0}(t + 1) = f[L(t); \Delta L(t)] \quad (6)$$

где:

$L_{class_0}(t + 1)$ – класс целевой переменной в момент времени $t + 1$;

$L(t)$ – целевая переменная в момент времени t ;

$\Delta L(t) = L(t) - L(t - 1)$ – изменение целевой переменной в период времени с $t - 1$ по t .

Далее будем сравнивать результаты, полученные нулевой моделью, с результатами более сложных моделей. Если они не улучшат точность прогнозирования по сравнению с нулевой моделью, то их применение является нецелесообразным.

3. Модель прогнозирования ликвидности с использованием сетевых моделей

Для целей построения модели будем использовать граф финансовых транзакций, который будет описывать движение денежных средств в течение одного дня как внутри банка, так и притоки, и оттоки от внешних контрагентов. Используются данные о денежных переводах за период с 2022 по 2023 год.

В качестве сетевых характеристик графа используются следующие показатели центральности узлов [4]:

- центральность по близости (Closeness);
- центральность по посредничеству (Betweenness);
- центральность по степени (Degree);
- центральность PageRank;
- центральность по Кацу (Katz centrality).

Будем использовать историческую информацию о целевой переменной и сетевых характеристиках графа финансовых транзакций для прогнозирования её значения при помощи алгоритмов машинного обучения. Полученный результат сравним с нулевой моделью и оценим, насколько велик вклад сетевых характеристик в повышении точности прогнозирования.

3.1. Построение агрегированного графа транзакций

Узлами графа финансовых транзакций являются: клиенты и контрагенты банка, сам банк, налоговые органы. Дугами – суммарные денежные переводы между узлами в течение дня [5]. Граф строится на данных за каждый день выборки (~500 графов за период с 2022 по 2023 год).

- Среднедневное количество узлов >1000, всего уникальных за период 2022-2023 год >10 000.
- Среднедневное количество дуг >5 000, всего за 2 года >2 млн шт.

Для упрощения расчетов будем в дальнейшем рассматривать агрегированный граф финансовых транзакций. Узлы графа агрегируются исходя из их принадлежности к одной из пяти групп: сам банк (*Bank*), клиенты банка (*Client*), дочерние компании банка (*Subsidiary*), банки-контрагенты (*Interbank*), налоговая служба (*Tax*). Дуга агрегированного графа представляет собой сумму денежных средств, перечисляемых лицами, принадлежащих одной группе, лицам, принадлежащим другой группе. Итоговый агрегированный граф имеет 5 узлов.

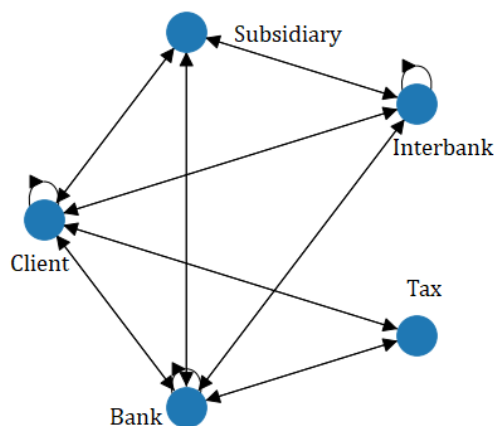


Рис. 2. Агрегированный граф финансовых транзакций

3.2. Модель прогнозирования

Для предсказания класса целевой переменной используются исторические значения самой целевой переменной, а также показатели центральности узлов агрегированного графа:

$$L_{class_n}(t+1) = f[L(t), \Delta L(t), C(t) \dots C(t-n), \Delta C(t) \dots \Delta C(t-n)] \quad (7)$$

где:

$L_{class_n}(t+1)$, – класс целевой переменной в момент времени $t+1$;

$L(t)$ – целевая переменная в момент времени t ;

$\Delta L(t) = L(t) - L(t-1)$ – изменение целевой переменной в период времени с $t-1$ по t ;

$C(t) = [\text{betweenness}(t), \text{closeness}(t), \text{degree}(t), \text{katz}(t), \text{pagerank}(t)]$ – матрица значений центральностей;
 $\Delta C(t) = C(t) - C(t - 1)$ – изменение центральностей в период времени с $t-1$ по t ;
 n – глубина исторической выборки.

4. Результаты

Для реализации алгоритмов машинного обучения разобьем исходные данные на три подвыборки в следующих пропорциях:

- обучающая (60% от выборки);
- валидационная (20% от выборки);
- тестовая (20% от выборки).

Так как речь идет о прогнозировании временного ряда, разбиение выборки осуществляется последовательно: сначала обучающая, затем валидационная и, наконец, тестовая. Выборки не перемешиваются.

Для реализации нулевой модели и модели прогнозирования ликвидности используем алгоритм машинного обучения – случайный лес – для задач классификации (RandomForestClassifier) [6].

Выбор меры центральности, глубины исторической выборки, а также гиперпараметров случайного леса осуществляется путем оптимизации показателя точности (*accuracy*) на валидационной выборке по следующему алгоритму:

```

1: Input:
   Вектор мер центральности  $c = [\text{betweenness}, \text{closeness}, \text{degree}, \text{katz}, \text{pagerank}]$ 
   Матрица значений мер центральности
        $C(t) = [\text{betweenness}(t), \text{closeness}(t), \text{degree}(t), \text{katz}(t), \text{pagerank}(t)]$ 
   Матрица объясняющих переменных
        $F(t, n) = f[L(t), \Delta L(t), C(t) \dots C(t - n), \Delta C(t) \dots \Delta C(t - n)]$ .
   Вектор целевых переменных  $L$ 
   Максимальное количество исторических дней для анализа  $n_{max}$ ,  $n \in [1 \dots n_{max}]$ ,
2: Output: Наилучшая мера центральности  $c_i$  и оптимальная глубина исторической выборки  $n$ 
   при  $accuracy(F_{val}, L_{val}) \rightarrow max$ 
3:  $Split(L) = L_{train}, L_{val}, L_{test}$ 
4:   for  $i$  in  $c$ :
5:      $F_i = f(C_i(t), n)$ 
6:     while  $len(\text{columns}(F_i)) > 3$  do:
7:        $a_0 = []$ 
8:       for  $j$  in  $1, \dots, len(\text{columns}(F_i))$ :
9:          $F_j = F_i \text{ del columns}(j)$ 
11:         $Split(F_j) = F_{train}, F_{val}, F_{test}$ 
12:        for  $d$  in  $1, \dots, depth_{max}$ :
13:          for  $e$  in  $1, \dots, est_{max}$ :
14:             $model \text{ fit} = \text{RandomForestClassifier}(F_{train}, L_{train}, d, e)$ 
15:             $a_0 \leftarrow \text{append } accuracy[model \text{ predict}(F_{val}), L_{val}]$ 
16:             $a \leftarrow \text{append } accuracy[model \text{ predict}(F_{val}), L_{val}]$ 
17:           $F_i = F_i \text{ del columns} \leftarrow \min(a_0)$ 
18:    $i, n \leftarrow \max(a)$ 
19: return  $i, n$ 

```

Рис. 3. Алгоритм определения оптимальных параметров модели

Оценка качества моделей выполнена на тестовой выборке. Сравнение результатов моделей приведено в таблице 3.

Таблица 3. Сравнение метрик качества нулевой модели (индекс - 0) и модели, использующей сетевые характеристики (индекс - n)

Метрика качества ¹	Class ₁	Class ₂	Class ₃
$Precision_n$	0.70	0.53	0.67
$Precision_0$	0.68	0.44	0.67
<hr/>			
$Recall_n$	0.61	0.43	0.90
$Recall_0$	0.68	0.44	0.67
<hr/>			
$Accuracy_n$	0.64		
$Accuracy_0$	0.59		

Точность прогнозирования целевой переменной исходя из ее исторических значений составляет 0.59. Использование сетевых характеристик графа транзакций повышает метрики качества модели: точность увеличивается до 0.64.

Мерой центральности, которая обеспечивает наилучшую точность модели, оказалась мера центральности по Кацу. Распределение важности объясняющих переменных приведено на диаграмме.



Рис. 4. Распределение по значимости объясняющих переменных

Несмотря на использование сетевых метрик, основной объясняющей переменной остается сама целевая переменная на предыдущем шаге - $L(t)$. Следующей по важности идет центральность узла, отвечающего за клиентские операции. Так как именно клиентские операции являются устойчивой базой для фондирования активных операций банка и, как видно из таблицы 1, составляют большую часть привлеченных средств банка.

5. Заключение

Применение центральностей графа транзакций в задаче прогнозирования ликвидности повышает точность и разделяющие способности модели. Показатель $accuracy$ повышается на 5 п.п. относительно нулевой модели.

Наилучшей центральностью для данной задачи оказалась центральность по Кацу, поскольку она учитывает веса связей и подходит для графов небольшого размера.

При этом точность классификации при равномерной выборке средняя. Улучшение модели возможно за счет:

- перехода от видов клиентов непосредственно к самим клиентам
- классификации видов операций, таких как оплата счетов, выдача/погашение кредита, размещение депозита, оплата налогов и т.п.
- учета иных параметров транзакций, таких как количество сделок, срочность операций, валюта и т.п.
- использования внешних переменных, не связанных с транзакционной активностью клиентов, например, процентных ставок, предлагаемых банком.

¹ Метрики качества [7]: $Precision = \frac{TP}{TP+FP}$; $Recall = \frac{TP}{TP+FN}$; $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
 где: TP - True Positive, FP - False Positive, FN - False Negative, TN - True Negative

Литература

1. *Банк России*, Инструкция №199-И «Об обязательных нормативах и надбавках к нормативам достаточности капитала банков с универсальной лицензией» от 29.11.2019.
2. *Банк России*, Доклад для общественных консультаций «Концепция макропруденциального стресс-тестирования», 2017.
3. *Блохин А.А., Стерник С.Г., Близняк А.Б., Телешев Г.В.* Институциональные различия российских банков по уровню доступа к фондированию путем сделок РЕПО // Дайджест Финансы. – 2023. – Т. 28, № 1. – С. 31 – 52.
4. *Newman M.E.J.* Networks: An Introduction, Oxford University Press, 2010. – 772 p.
5. *Akrati Saxena, Yulong Pei, Jan Veldsink, Werner van Ipenburg, George Fletcher, Mykola Pechenizkiy*, The Banking Transactions Dataset and its Comparative Analysis with Scale-free Networks // Eindhoven University of Technology, 2021.
6. *Breiman L.* Random Forests. // Machine Learning. 2001;45:5–32.
7. *Sammut, C., Webb G.* Encyclopedia of Machine Learning. Springer, 2011. – 1031 p.