第3章 Zero-shot/Few-shotプロンプティングを使った異常検知

実験1…手軽に異常を判断する方式を試す

氏森 充

表1 Zero-shot検知とFew-shot検知の比較

方 式	手 法	利 点	欠 点	推奨シーン
Zero-shot検知	Zero-shotプロンプティングを使用(プロンプトを事例なしで，指示文のみで検知)	• 事例なしで即座に検知可能 • 未知の攻撃パターンにも対応可能	• 精度が低い場合がある • 誤検知が多い可能性がある • ドメイン固有の知識を活用できない	• 新しいシステムや未知の脅威 • 概要把握や初期調査
Few-shot検知	Few-shotプロンプティングを使用(プロンプト中に事例と指示文を含めて検知)	• 精度が大幅に向上 • 誤検知率が低い • ドメイン固有の知識を活用	• 適切な例が必要 • セットアップに時間がかかる	• 既知のパターンがある場合 • 高精度が必要な場合

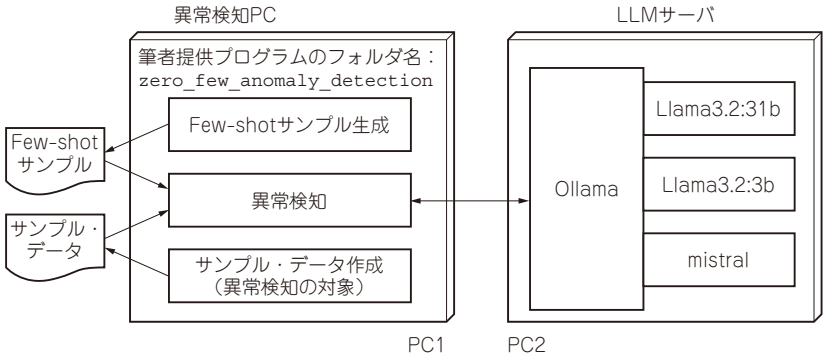

図1 ローカルLLMを使った異常検知，実験のシステム構成

第3章，第4章では，ローカル環境で動作する大規模言語モデル(ローカルLLM)を活用し，ネットワーク・トラフィック，システム・ログ，センサ・データなど，多様な入力データから異常を検知する方法とプログラムを解説します．プログラムは，Ollamaを用いたローカル実行型のLLMを利用します．

扱う手法

● Zero-shot検知

Zero-shot検知は，プロンプト中に例示を与えずに異常を判断する方式です．LLMが持つ言語理解能力と一般的な知識に基づき，プロンプトに記した指示内容と入力データの内容から，

- 通常と異なる振る舞い
- 矛盾した状態

を推論して検知する点が特徴です．

一方で，事前の学習が全く行われていないため，検知精度が安定しにくいという課題もあります．未知のシステムや初期段階の分析などで，大まかに傾向をつかみたい場合に適しています．

● Few-shot検知

Few-shot検知では，プロンプト中に正常および異常の典型的な少数サンプルをあらかじめ提示します．これにより，LLMが提示された文脈を理解した上で検知を行えるため，単純なキーワード・マッチやしきい値判定では見逃されるような微妙な異常パターンも検出できます．

Zero-shot検知以上に，サンプルを基にしてなぜ異常と判断されたのかという理由付けを伴う説明的な異常検知を行える点が大きな特徴です．

各検知の特徴を**表1**に示します．