**Mapping:**

1. **Txn -----🡪 word**

**Amt, txnType, receOrpay, chipsOrfed, interbankAmt, JurisdictionAmt, swiftType, etc. time stamp**

**‘Amt\_ txnType\_ receOrpay\_ chipsOrfed\_ interbankAmt\_ JurisdictionAmt\_ swiftType’**

1. **Txn under a single customer -----🡪doc**
2. **Txn group by all customers---🡪all docs--🡪corpus**

**Issue: Why not use structured data but sequence text data?**

**Pre-training downstream**

**Embedding:**

**Pre-train data is from FinCEN**

**Frame: W2V, ELMO (bi-direction LSTM), GPT (transformer), BERT (bi-direction, transformer)**

**Above frame is good for detecting sequence pattern which is useful for customer suspicious behavior detection.**

1. **Customer segementaion**

**Txn---🡪(1.0,2.2,3.4)--🡪f(data,dim=3\*1)**

**Txn--🡪balckbox-🡪susp-🡪why? (tagging)**

**23 rule-based scenarios (typology manytoone, onetomany) --🡪file to TM team. False positive rate 98%**

**BERT LSTM-🡪 attention**

**Sigmoid**

**Dense fine-tune**

**Dense fine-tune**

**Dense fine-tune**

**Dense fine-tune**

**Attention fine-tune**

**Time-distribute fixed-par**

**LSTM**

**Embadding**

**Wire\_txn(swift message, customer CDD information)**

**---🡪 Wire Room (check the format / send RFI**

**and detect the suspicious txn / directly send to TM for investigation) (non-auto: this paper focus on here)**

**---🡪 OFAC screening (auto)**

**---🡪 AML system (auto)**

**---🡪 TM (if txn be alerted by scenarios then will be sent to TM for investigation) (rely on human experience: future work)**

**--🡪 SAR (if customer is determined as true positive, will be filed to FinCEN) (auto)**

**The automation ratio is 40%**

**Future Work: Customer Seg+ Susp Detection**

**Customer Seg:**

**1.NLP approach:**

1. **Hierarchical Dirichlet Process**
2. **NMF (Matrix Decomposition)**
3. **Semi-LDA, Semi- Hierarchical Dirichlet Process**

**2.Normal structure data:**

1. **Clustering, Semi-clustering**

**Susp Detection:**

1. **Multivariate time series Anomaly detection**
2. **NLP Anomaly detection**
3. **NLP text similarity**
4. **GAN can be helpful**