

Line Developer day 2019 報告

- Key Note (資料)
- シンプルかつ効果的なデータ拡大により強力かつ堅固な深層モデルを作る (資料)
- Inside of Blog: 15年熟成されたサービスの光と影、カオスとレガシーへの挑戦 (資料)
 - 次の15年後
- LINE NEWSの記事配信を支える技術 (資料)
 - ▶レコメンドされたニュース
 - ▶マシンラーニング側のシステムフロー図
 - ▶Redisのレコメンドデータを保存仕組の問題(特に手動登録)
 - ▶これからは、
- コミュニケーションアプリ「LINE」の機能改善を支えるデータサイエンス (資料)
- Reliability Engineering Behind The Most Trusted Kafka Platform (資料)



事前に顔写真を登録して、顔認証で会場に入ったが、精度が高く反応速度も速くてびっくりした！

Key Note (資料)

- 毎日1兆件の新しいデータが生まれていて、390 TBのデータボリューム。
- 今まで、アカウントをハッキングして乗っ取る / キャンペーンの際にお金目当てで組織的にアカウントを複数作られるなどの問題がありましたが、マシンラーニングを使い、何年間戦った結果、2018年で0件を記録した。
- 台湾のLineでは、ニュースが本当かどうか判別する機能を発表したらしい。
- AI CallといったAI電話対応サービスを開発していて、現在ベータ版を一つのお店と連携し検証中である。(関連記事)
- これが実用化されると、電話対応が多い部署に使えるのかもしれない！
- レポートを手書きで書くことに不満を持っていた大学生が全自動手書きマシンを作って、それが話題になった (関連記事)
- その大学生を招待し、マシンの動く姿を披露してもらった。

シンプルかつ効果的なデータ拡大により強力かつ堅固な深層モデルを作る (資料)

- AIの写真認識についての話
 - 画像の認識精度を高めるために、いろいろなパターンの画像を用意して努力してきた
1. 回転した画像
 2. 部分的に取りぬいた画像 (Cut Out)

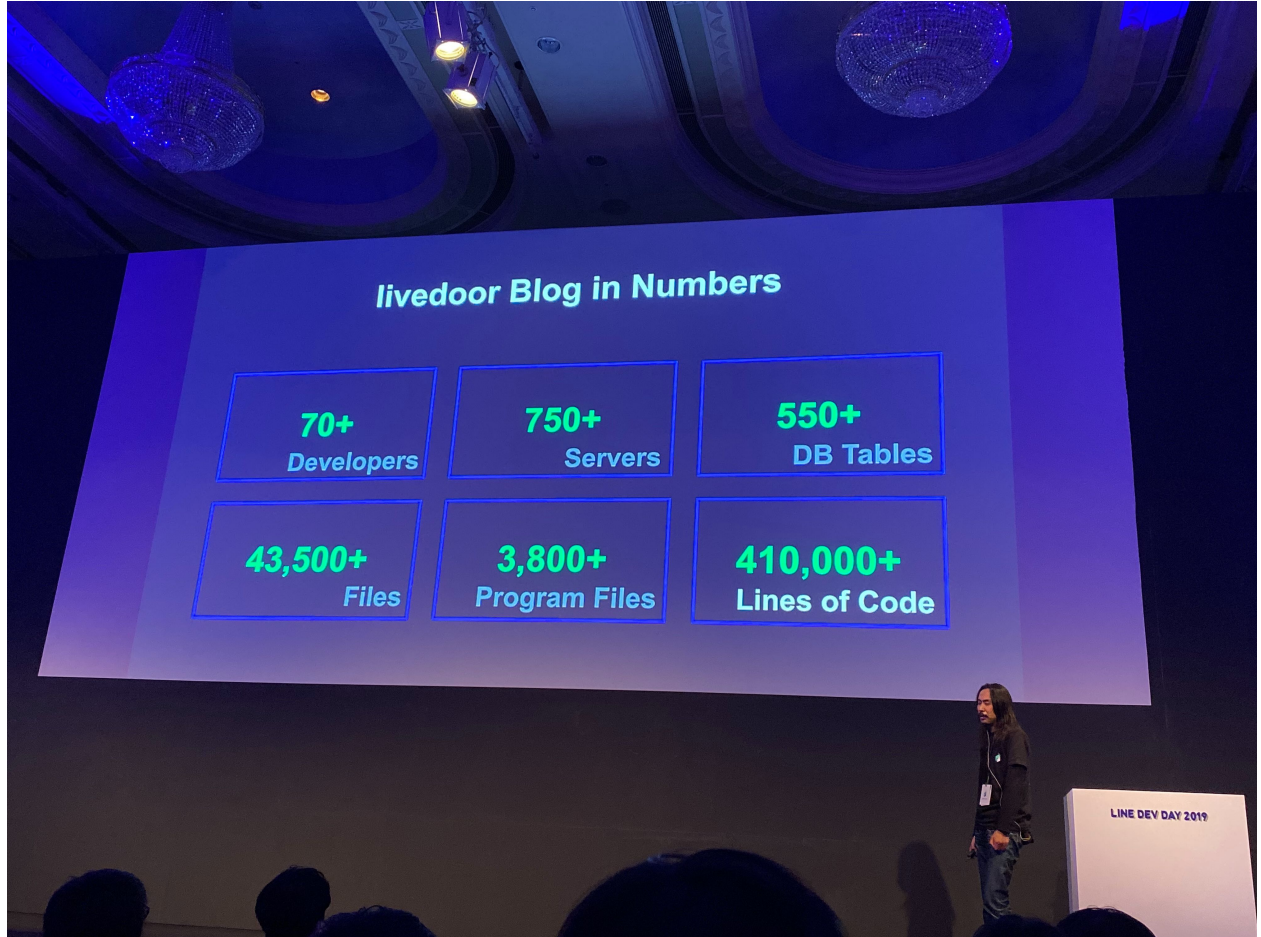
3. 他の写真とレイヤーを重ねた画像 (Mix Up)

4. 他の写真と合成した画像 (Cut Mix)

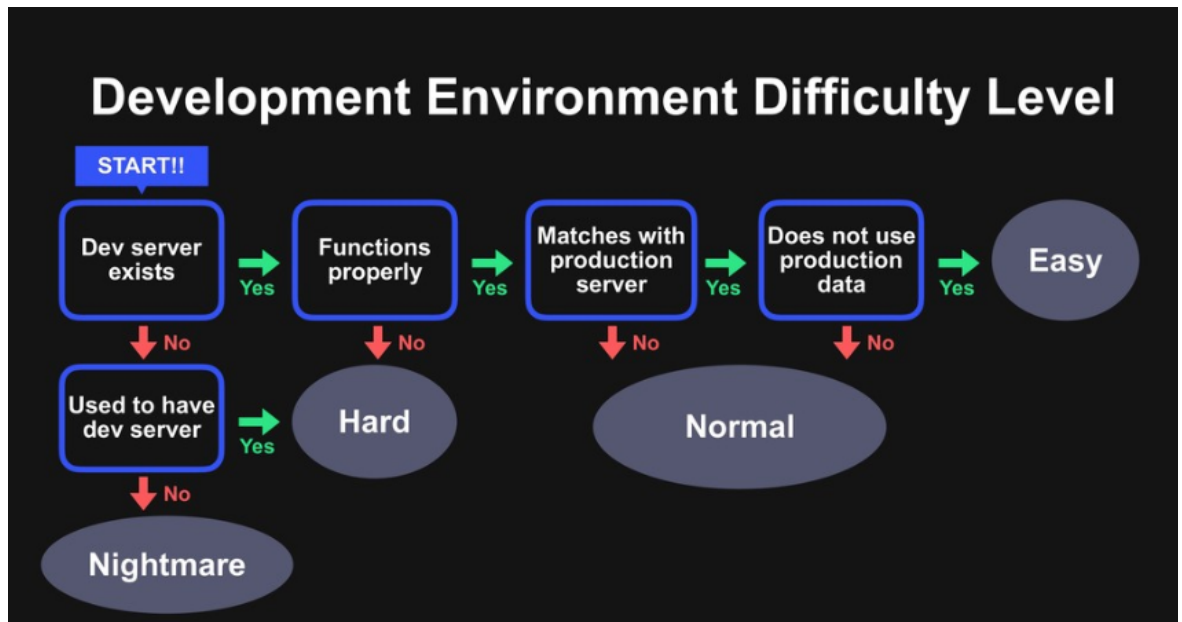
- 4のCut Mixのパターンでは、どんなものが、どこに、どんな大きさに混ざっているのかを判別しなければならないので、大変だった HitMapの比較を行った!

Inside of Blog; 15年熟成されたサービスの光と影、カオスとレガシーへの挑戦 (資料)

- Livedoor Blogの話



- LAMP環境 (PはPHPではなくPerl) Eストアーと似た感じ!
- どんな問題があったか
 - ドキュメントがすぐ見つからない
 - 本番と似た開発環境が見つからない



- テストされていないコードが多い
- 不要なDNS Recordsが多い
- 機能がすぎる
- 言語のバージョンが低い (Perl 5.8, **MySQL 4.0** :scream:)



- これから時間をかけて、レガシーなところを変えていこうとする計画を立てているらしい。



次の15年後

最新の技術や開発方式を行っても15年後にはレガシになるので、現在開発する以下のように開発を行う。

小さい改善からスタートする

後回ししないで、今すぐ行う

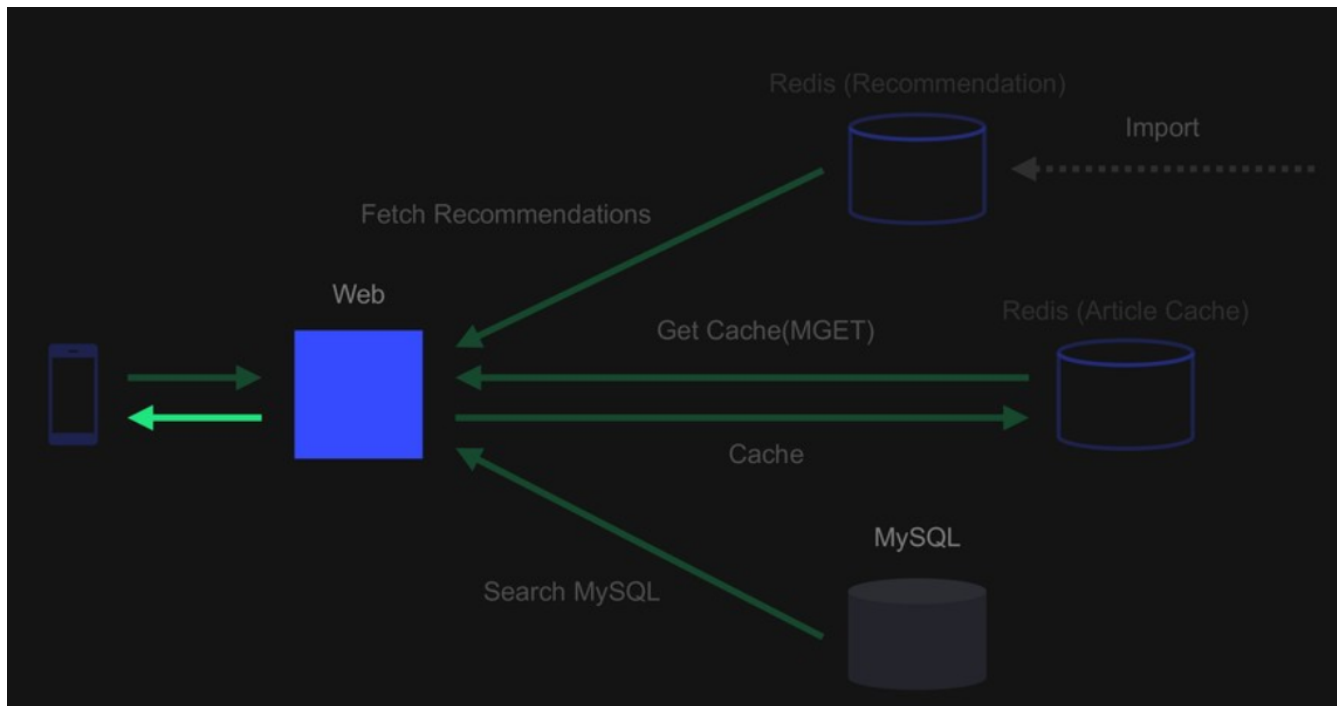
いつも意識を持ってコードを書く

LINE NEWSの記事配信を支える技術(資料)

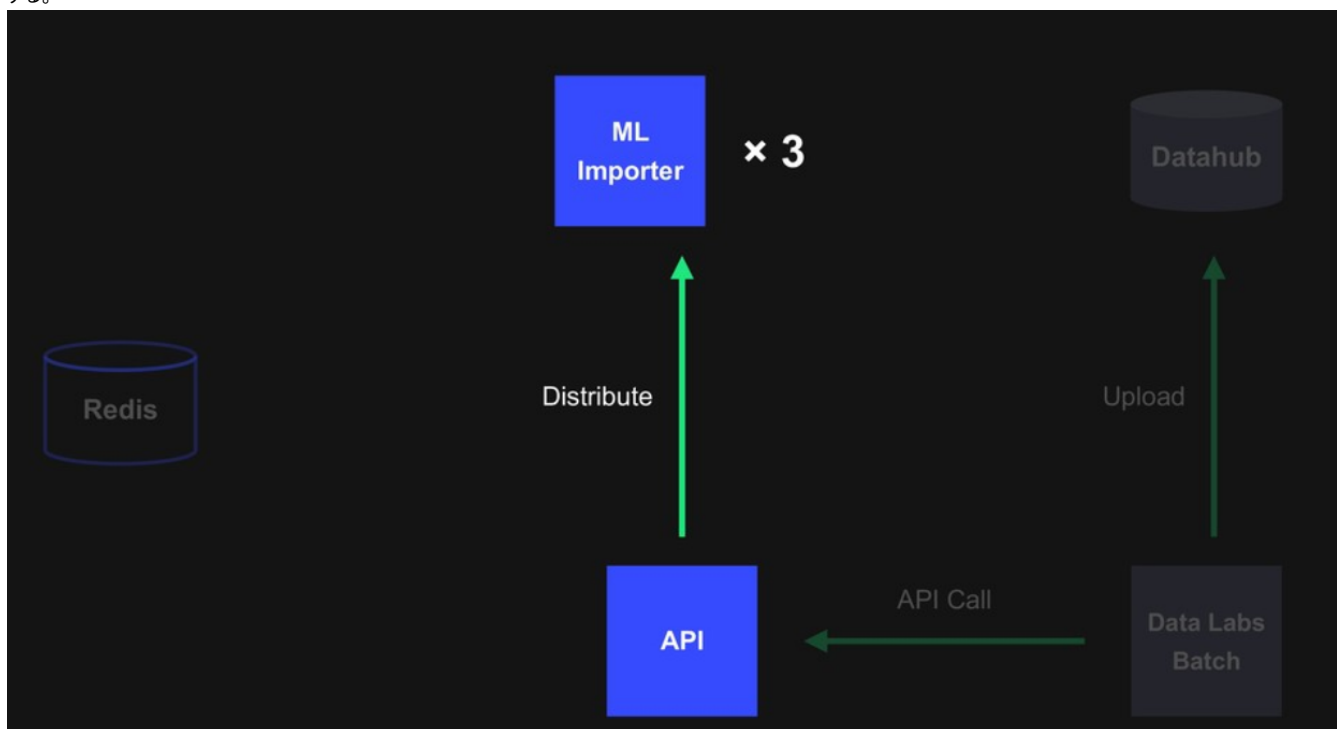
▶レコメンドされたニュース

ユーザーに合わせたニュースを見せるため、マシンラーニング+手動対応を行ってる
1時間ごとに提供するデータ: 100万ユーザー * 200ニュース = 2億
最初マシンラーニングだけで提供を行いましたが、
年齢や性別、地域のような条件の組み合わせはオペレーターが行う。

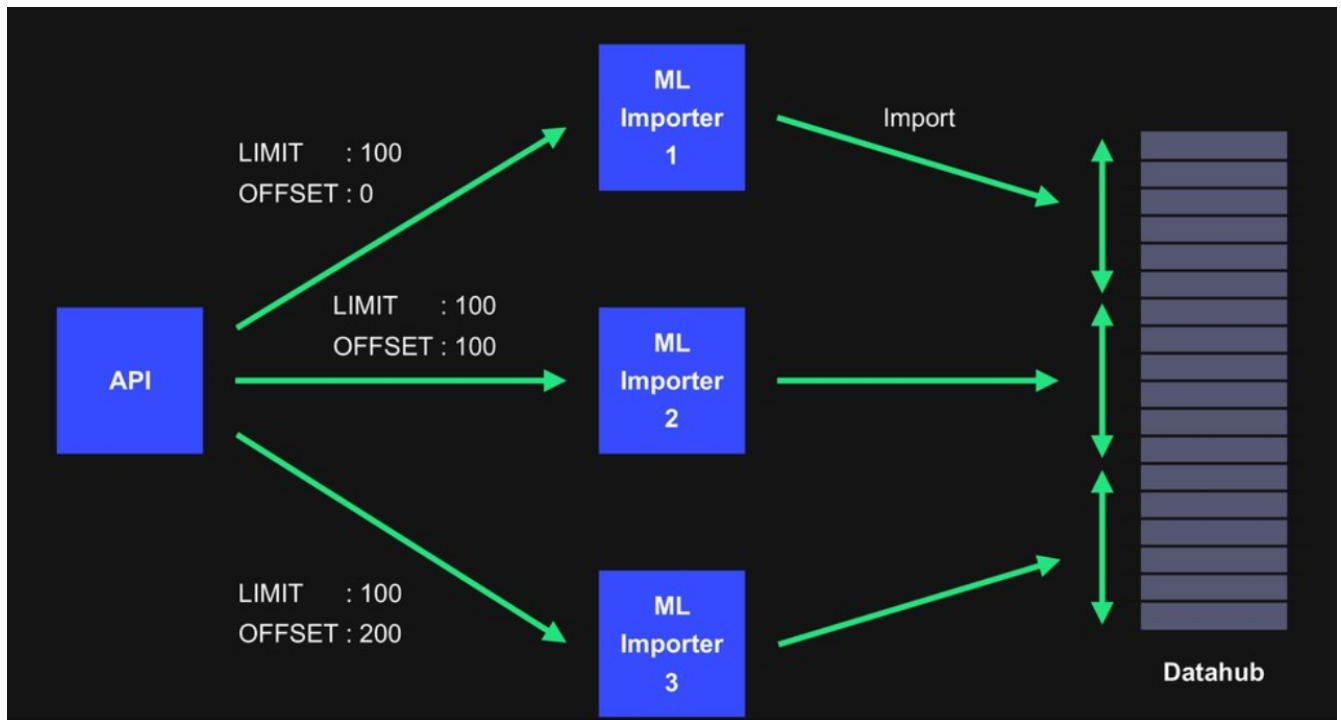
▶マシンラーニング側のシステムフロー図



まず、mysqlで記事やユーザの情報を保存し、レコメンドデータやキャッシュはRedisを利用する。
 ただ1時間ごとに新しい200件の記事に対する100万ユーザーをレコメンドデータに入れる必要があったので、レコメンドデータを入れる仕組みを改善する。



- ① データをDataLabsBatchでDatahubにアップロードし、APIを呼ぶ。
- ② APIは3つのマシンラーニングサーバでマシンラーニングを行い、DataHubに保存する。



③ Datahubからファイルをparseして、Redisに保存

▶Redisのレコメンドデータを保存仕組の問題(特に手動登録)



記事ごとにユーザーを設定するため、(記事番号、ユーザー番号で探した場合)

1. 拡張性が低い
2. 登録コストが高い

そのため、下部の仕組に変更で、ユーザーで好みの記事を探せるようになる。

▶これからは、

1時間ごとの更新ではなく、リアルタイムにしたい
他のパーツでもレコメンドを投入したい

コミュニケーションアプリ「LINE」の機能改善を支えるデータサイエンス(資料)

- 話題は、グループ開設機能の改善。
- グループというのは、人の社会的人間関係を表している。
- グループ開設する途中で離脱してしまう問題を解決したい!
- 現状:グループアイコン・名前を決める メンバーを決める
- :point_down:色々なソリューションを実装して、ログ分析を行い、効果測定を行った(=データサイエンスチームの仕事)
- ソリューション1

グループ名・アイコンを決めづらいから、離脱するんだ！

メンバーを招待してから、グループ名・アイコンを決めさせる

しかし、効果ない・・・

- ソリューション2

招待するメンバーをスムーズに選ばないから、離脱するんだ！

最近連絡した友達リストを上部に配置

これも、そんなに効果なし

- ソリューション3

離脱率はおいというて、そもそもグループを作れる窓口が少ないのではないかな

ホームタブ / 友達追加のメニュー / チャットを始めるメニューでグループを作れるようにする！

グループを作る人数が増えた！！

Reliability Engineering Behind The Most Trusted Kafka Platform (資料)

- Kafka platformは、LINEやLINEの運用型広告プラットフォーム、LINE Payなど特に高いサービスレベルが要求される多くのサービスで使われている。そのため、サービスとしてのレベルが非常に高い。
- 応答時間を10ミリ秒以下に抑えながら、クラスターの可用性を100%に近い状態で維持するとともに、1日当たり3,600億件以上のメッセージという膨大なトラフィックを扱う。
- システムの安定性と高いパフォーマンスを維持する上でやっている。信頼性を高くするため、以下の方針を持つてる。

原因を把握せずに、対応だけ行わないこと。

しっかり原因を把握した上で適切な対応を行う。

- その結果、1年内のサーバ周りの通常率が 99.9%~100%
- 例) クラスターサーバが落ちてしまう(JVM Pause Time Increase)

JVMのエラーについて勉強

JVM threadが一時的に中断するまでに待つ

Diskアクセスするのに時間が増加している。

リナックスでfile I/O測定する速度を測定

HDDが壊れてる事情を発見

別のサーバ運用チームでrebootすることで再現が出来なくなる

リナックスコアーで無理有り、Disk I/Oを遅くする。

kafkaをエラーにさせるため、fileを無限に書くCを作成

再現成功 → 適切な解決方法を発見

- 難しくて、よく分からないところが沢山ありました ;(