

公開シンポジウム
「事故による子どもの傷害を予防する
ー子ども中心の新たな予防システムの構築へ」

表データを利活用するためのAI技術

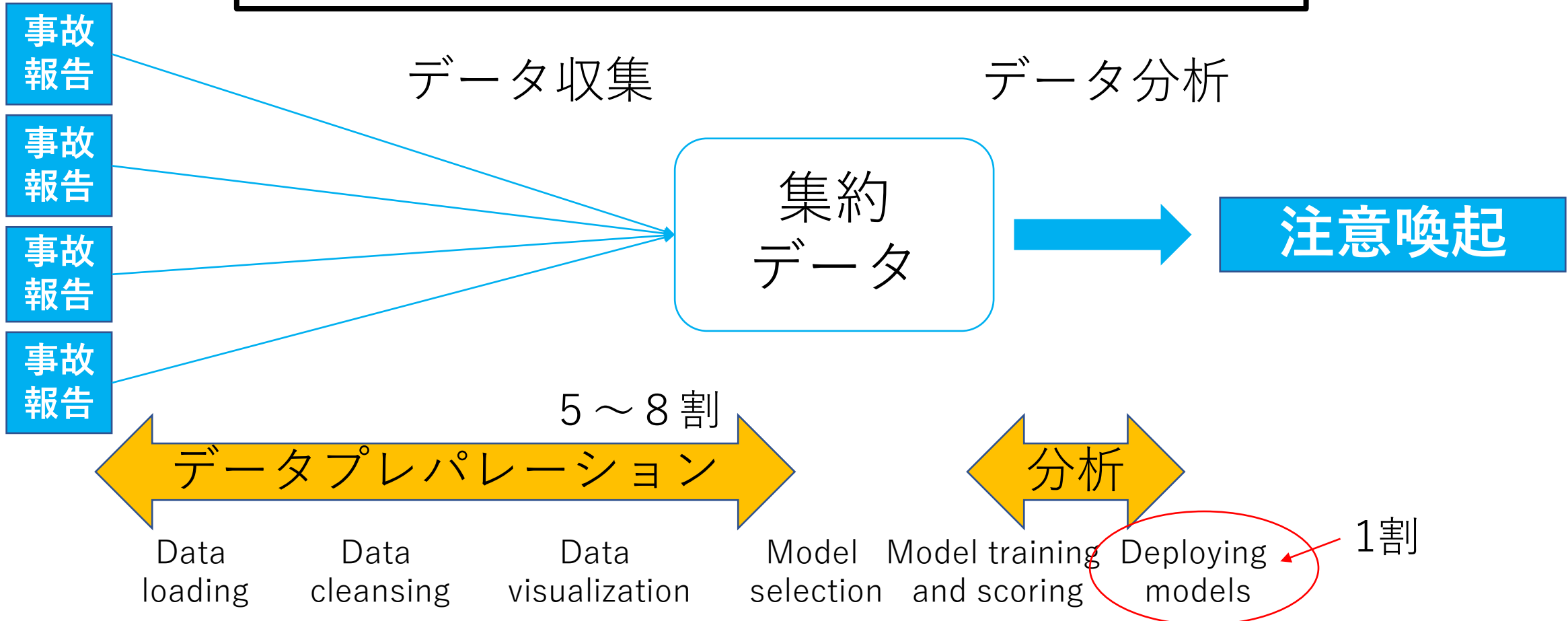
相澤 彰子

日本学術会議第三部会員

国立情報学研究所教授

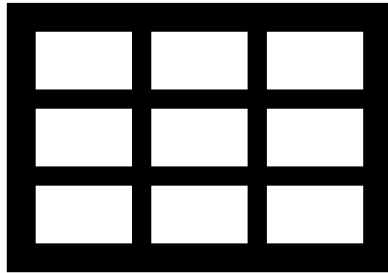
データ収集・データ分析

労力の多くはデータの準備に費やされる



「表データ」の利活用

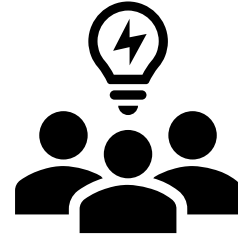
表データ



ユーザ



データ検索



入力データの質を高める

- 入力項目の適切な設定
 - 語彙やフォーマットを決める
 - 使い勝手のよい入力フォームを提供
- 高度な検索機能
- タグ付与
 - 類似事項検索

人間のための「データ」可視化



AI

データプレパレーションの
ほんの入り口…

データ連携



テーブルを統計処理可能な形に整形する

- 欠損値を補う
 - 自由記述欄の解析
- 複数のテーブルをつなぐ
- 項目マッピング
 - 値のフォーマットを揃える
 - 項目の統合、分割

AIの活用例

Wikipediaなどの大量の文書でモデルを訓練して、事前に知識を埋め込む

表と知識グラフの対応付け

固有表現抽出

エンティティ
リンキング

関係抽出

深層学習 (AI)による性能改善



類似検索

夫が生後4ヶ月の乳児をプ्रेसイミング用フロートを使用して入浴させたところ、窒息し呼吸停止で入院。

浴槽で、赤ちゃん用の首浮き輪から外れ、心肺停止状態になり救急搬送された。

Premise

浴槽で、赤ちゃん用の首浮き輪から外れ、心肺停止状態になり救急搬送された。

Hypothesis

赤ちゃんが浴槽で溺れた。

Entailment

Yes/No/Neutral

言語推論

言語モデルに推論機能を学習させる

潜在トピック解析

表の構造認識