PresenSum: トーク音声とスライド画像情報を入力とするプレゼンテーション要約

斉藤いつみ　　西田京介　　吉田仙
日本電信電話株式会社　NTT 人間情報研究所
{itsumi.saito.df, kyosuke.nishida.rx, sen.yoshida.tu}@hco.ntt.co.jp

概要

プレゼンテーションのトーク音声とスライド画像集合を入力として，テキストの要約を出力する新たなプレゼンテーション要約タスク PresenSum を提案する．学会発表の共有サイトからデータセットを収集，分析し，本タスクが既存のテキスト要約タスクと異なる特徴があることを確認した．さらに，テキストで事前学習された Encoder-Decoder モデルをベースとしたプレゼンテーション要約モデルを構築し，トーク音声の音声認識テキストとスライド画像の文字認識テキストの両方を入力することの効果やスライド中の単語の大きさ，配置などのレイアウト情報や画像特徴などを追加する効果を検証した．

1　はじめに

テキスト要約技術は言語処理における重要な技術分野の一つであり [1]，言語モデルの進展とともに大きく発展している [2, 3, 4]．一方，現実社会における様々なニーズに対応するためには，テキスト以外の音声や画像などの情報も入力とするマルチメディアデータの要約も重要である [5, 6, 7]．本研究では，プレゼンテーションのトーク音声とスライド PDF ファイルから得られた画像情報を入力とし，テキスト要約を生成する新たなプレゼンテーション要約タスク PresenSum を提案する．プレゼンテーションを対象とする要約研究は一部存在するが，トーク音声のみを入力とするタスク [8, 9] や，動画中の粗い画像情報を利用するタスク [10] であり，高精細なスライド画像情報とトーク音声を同時に考慮したプレゼンテーション要約については取り組まれていない．

図 1 に PresenSum の概要を示す．本研究では，トーク音声とスライド画像から抽出したテキスト情報に加え，スライドテキストのレイアウト（テキストのサイズや配置情報）や画像情報を利用した要約

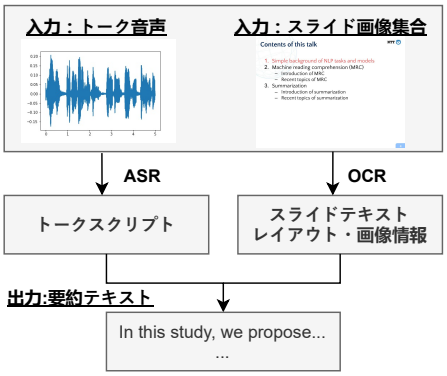

図 1　PresenSum タスクの概要．トーク音声，スライド画像集合を入力してテキスト要約を生成する．

生成を検証する．音声認識によって得られたトークスクリプトは誤りも含むが，スライドテキストにはキーワード的に重要な内容が記載されているため，双方を入力することで要約の質が向上することが期待できる．また，スライド中の文字には色や大きさ，配置などで重要箇所が視覚的に理解しやすい装飾が施されているため，これらの情報も要約において重要な情報と考えられる．PDF などの文書画像の質問応答に関する研究は最近盛んに行われており [11, 12, 13]，文書画像中の単語の配置や画像情報の有効性が示されているが，スライド画像を含む要約タスクにおいて有効性は検証されていない．

本研究の貢献は，1) PresenSum のデータセット収集と分析，要約モデルを構築し，2) トークスクリプトとスライドテキストの双方を入力することの効果検証，3) スライドテキストのレイアウト・画像情報を入力することの効果検証を行った点である．

2　PresenSum データセット

2.1　データセットの収集

学会等のプレゼンテーション動画・資料がアーカイブされている videolectures.NET[1)] から動画, スライ

1）　http://videolectures.net/

表 1 既存の要約データセットとの比較.

Dataset	単語数 入力	単語数 要約	Coverage	Density	Comp
CNN/DM	776	53	0.823	2.94	14.9
XSum	454	24	0.665	1.09	19.8
ArXiv	6,322	289	0.881	3.42	40.2
PubMed	3,202	213	0.853	4.79	16.1
PresenSum	5,611	187	0.823	1.95	37.2
ASR only	3,043	187	0.748	1.34	20.3
OCR only	2,568	187	0.699	1.52	16.9

ド PDF データ, 要約テキストが存在しているデータを収集した．要約テキストは，上記サイトの各プレゼンテーションページに記載された Description テキストを収集した．動画から音声データを抽出し，スライド PDF はページごとに画像に変換した．データ数合計は 2305 件であり，動画の平均時間は 1,100 秒，スライドの平均枚数は 30.8 であった．トークスクリプト，スライドテキストはそれぞれ Google Speech/Vision API[2] を用いて抽出した．以降，それぞれ ASR テキスト，OCR テキストと表記する．

2.2 既存の要約データセットとの比較

本タスクは入力にテキスト以外の情報も含むが，ASR や OCR によって変換したテキスト情報が代表的な入力となるため，本節ではテキスト情報のみに着目した比較を行う．表 1 に代表的なテキスト要約データセットである CNN/DM [14]， XSum [15]（ニュースドメイン）， ArXiv [16]， PubMed [16]（論文ドメイン）と PresenSum の比較を示す．既存データセットについてはそれぞれ train データから 10000 件をランダムサンプリングして指標を計算した．

比較指標 各データの特徴を表す指標として，入力・要約の単語数と，[17] にて提案された Coverage, Density， Compression の 3 指標の合計 5 つの指標を用いる．Coverage は要約の単語が入力テキストによってどの程度カバーされているか，Density は要約と入力テキストが連続して一致した部分単語列の平均長がどの程度か，Compression は入力テキストが要約テキストの何倍の長さかを表す．

入力・要約単語数の特徴 PresenSum は，ASR と OCR を合計した入力単語数が平均 5,611 語，要約単語数が平均 187 語と，ニュース要約（CNN/DM, XSum）に比べて入出力が長い．入出力単語数の観点では Compression 指標も含め論文要約タスクの ArXiv に近いと言える．

2） https://cloud.google.com/{speech-to-text,vision}/

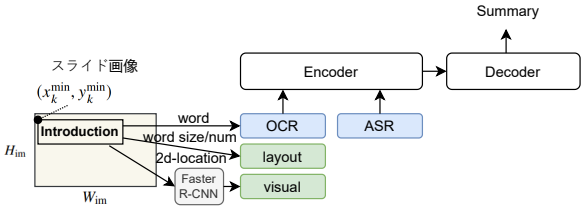

図 2 プレゼンテーション要約モデルの構成．Encoder に OCR および ASR テキストと，OCR テキストのレイアウト情報や画像情報を追加する．

入力・要約単語の一致度の特徴 PresenSum の Coverage は CNN/DM や ArXiv, PubMed と同程度の高い値であり，入力テキストの単語を適切に利用しながら要約を生成することが求められるタスクと言える．また，ASR/OCR 単体のデータに比べて PresenSum データでは Coverage が大きく向上しており，ASR/OCR テキスト双方を入力することで要約に必要な単語が補完されることが示唆される．一方，Density について PresenSum は Arxiv 等に比べて低い値である．これは，要約テキストに出現する単語が ArXiv 等と比べて入力テキスト中に散らばって存在しており，連続した単語列の表現をそのまま利用する割合が少ないことを示す．ASR テキストは話し言葉であり冗長表現や認識誤りも含まれること，OCR テキストもキーワード的な表現が多く要約テキストとスタイルが異なることが原因と考える．

以上の指標から，PresenSum は正解要約の単語が入力テキスト中にどのように分布しているかについて従来の要約データとは異なる特徴を持ち，Compression が高く Density が低い点で既存データに比べて要約生成が難しいデータセットといえる．

3 プレゼンテーション要約モデル

図 2 にモデル構成を示す．生成型要約を行うため，Encoder-Decoder モデルをベースモデルとした．ASR，OCR テキスト系列をベース入力とし，レイアウト，画像特徴量系列を追加する．

テキスト系列 ASR テキスト単語系列 X_{asr}, OCR テキスト単語系列 X_{ocr} を利用する．これらのテキストをモデルに対応したトークナイザでサブワードに分割し，OCR, ASR を表す特殊トークンを用いて [OCR] + X_{ocr} + [ASR] + X_{asr} のように結合する．

レイアウト系列 OCR によって検出された単語の配置情報 X_{loc}，単語のサイズ情報 X_{size}，各単語が属する段落中の単語数情報 X_{num} を利用する．配置情報は $x_k^{loc} = [x_k^{\min}/W_{\text{im}}, y_k^{\min}/H_{\text{im}}, x_k^{\max}/W_{\text{im}}, y_k^{\max}/H_{\text{im}}]$

をそれぞれ 1000 倍した整数 [18] とする．サイズ情報 x_k^{size} は各単語の高さ情報 $y_k^h = y_k^{\max} - y_k^{\min}$ を計算し，各プレゼン内の平均サイズ y_{ave}^h を用いて $y_k^{h'} = y_k^h / y_{ave}^h$ としたあと，10 倍して整数化した値を用いる．なお，$(x_k^{\min}, y_k^{\min})$，$(x_k^{\max}, y_k^{\max})$ はトークンを囲む矩形領域の左上および右下の座標，W_{im}，H_{im} は画像の幅及び高さを表す．

画像特徴量系列　OCR 単語の矩形領域ごとに COCO [19] で学習された Faster-RCNN [20] を用いて box head の最終層のベクトルを抽出し画像特徴量系列 E^{vis} として利用する．

Encoder 入力 Embedding　テキスト・レイアウト・画像特徴量系列を embedding し，次のように結合する．

$$e_t = \text{LN}(e_t^{\text{text}} + e_t^{\text{add}})$$

LN は Layer Normalization [21] を示す．e_t^{text} は，OCR トークン x_t^{ocr} あるいは ASR トークン x_t^{asr} を基に埋め込む．e_t^{add} は e_t^{loc}，e_t^{size}，e_t^{num}，e_t^{vis} のいずれかを表す．e_k^{loc}，e_k^{size}，e_k^{num}，e_k^{vis} は対応するレイアウトトークン，画像特徴量を線形変換にて埋め込む．ASR トークン部分の e_t^{add} については零ベクトルとする．

Encoder　Encoder は L 層の Transformer とする．Encoder 入力 Embedding 系列 $H^0 = [e_1^0, ..., e_T^0]$ を受け取り，$H^L = [e_1^L, ..., e_T^L]$ を出力して Decoder に渡す．

Decoder　Decoder は L 層の Transformer を用いて，通常の Text-to-Text と同様に生成したトークンを 1 単語ずつ入力し次のトークンを予測する．Decoder では入力にテキスト以外の情報は入力しない．

4　評価実験

評価データ　PresenSum データセットを 8:1:1 の割合でランダムに分割し，訓練，開発，テストデータとした．各データ数は 1844，230，231 である．

評価指標　ROUGE [22] を使用した．また，正解要約中の単語が生成テキストに含まれるか否かを 1/0 として品詞ごとに集計した品詞カバー率を評価する．品詞付与には spacy[3] を使用した．

実験設定　最長で 16384 トークンを扱うことができる事前学習済の Longformer Encoder-Decoder モデル [23] を使用し，入力系列が上限を上回る場合は先頭から用いた．学習済モデルの取得や学習は Transformers[4] を使用した．バッチサイズ 8，勾配の累積ステップ数を 2，10 エポック学習し，開発セットでの損失が最小のモデルを選択した．最適化手法は AdamW [24]，学習率は 5e-5 とした．

3）　https://spacy.io/
4）　https://github.com/huggingface/transformers

表 2　ASR テキストと OCR テキストの双方を入力する効果．カッコ内の数字は入力トークンの最大長を表す．本評価ではレイアウト・画像特徴量系列は使用しない．

Input	R-1	R-2	R-L
ASR (4096)	38.78	9.78	35.19
OCR (4096)	39.39	9.87	35.73
ASR+OCR (4096)	40.55	10.77	36.81
ASR+OCR (8192)	**40.96**	**11.09**	**37.23**

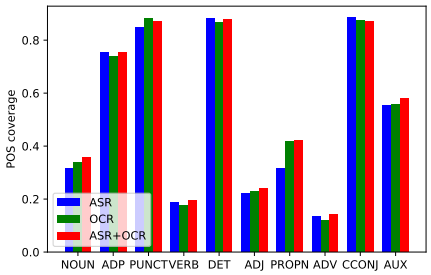

図 3　生成された要約が正解要約の単語を含む割合．

4.1　評価結果

ASR テキストと OCR テキストの双方を入力することは効果があるか？　表 2 に示す通り，ASR テキストと OCR テキストの双方を入力した場合が ASR，OCR 単体を入力する場合よりも精度が良かった．また，OCR 単体の方が ASR 単体よりも精度が高かった．2.2 節の単語の Coverage 指標では ASR 単体の方が OCR 単体よりも高い値だったが，要約生成の結果では逆になった．ASR の認識誤りや話し言葉のスタイルが精度に影響していると考える．

入力長を伸ばすことにより精度は向上するか？　表 2 に示す通り，ASR+OCR の場合の入力の最大長を 4096 から 8192 に長くすると精度が向上した．PresenSum は入力系列全体に分散した情報を要約することが必要なデータセットと言える．

ASR テキストと OCR テキストはどのような補完関係があるか？　図 3 に出現頻度が高い 10 種の品詞ごとの品詞カバー率を ASR，OCR，ASR+OCR の入力ごとに示した．まず，OCR を入力することで固有名詞（PROPN）のカバー率が大きく向上した．ASR においては固有名詞の認識が難しいが，OCR では固有名詞が取得しやすいため特に補完効果があったと考える．最も出現頻度が高い普通名詞（NOUN）においても ASR と OCR の双方を入力することでカバー率が向上した．

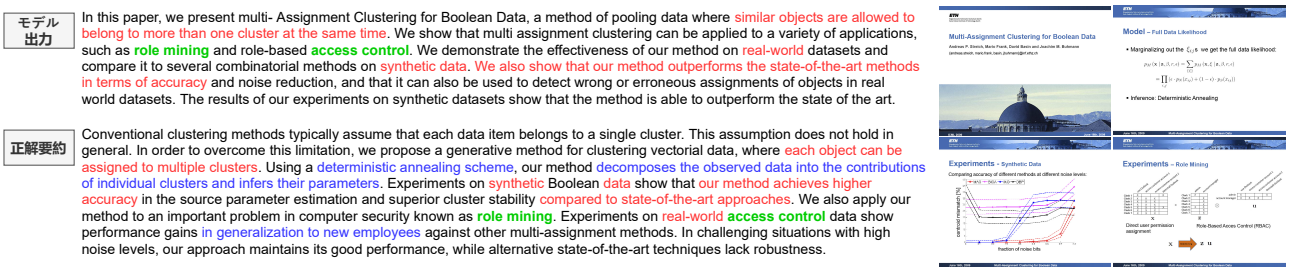

図 4　ASR と OCR の双方を入力した場合の要約出力例．赤文字が ASR や OCR 単体入力においても生成された内容．緑文字が ASR と OCR を両方入れた時のみ生成された内容．青文字はすべての要約で生成されなかった内容．

表 3　OCR テキストのレイアウト・画像情報の効果．

Input	R-1	R-2	R-L
ASR+OCR (4096)	40.55	10.77	36.81
+サイズ	40.60	10.93	36.95
+段落中単語数	**40.81**	**11.03**	**37.03**
+配置	40.17	10.55	36.39
+画像情報	40.40	10.95	36.72

OCR テキストのレイアウト・画像情報の効果はあるか？　表 3 に OCR のレイアウト情報や画像情報を追加した場合の効果について検証した結果を示す．OCR のサイズ（e^{size}），段落中の単語数（e^{num}）を追加するとベースラインよりも精度がわずかに向上した．段落中の単語数は単語のまとまりを表し，本文中のテキストであれば多くなり，図表中の語であれば 1 語で独立しているケースが多いなど，OCR テキストがどのような領域に属しているかを間接的に表す．このような OCR テキストの意味的な領域の理解をさらに深めることが重要と考える．
一方，配置情報（e^{loc}），画像情報（e^{vis}）については精度が向上しなかった．ベースとした Longformer Encoder-Decoder が大量の言語データで学習されている一方，今回の学習データが少量のため追加特徴量への適応が不十分であったと考える．

出力例・エラー分析　図 4 に ASR と OCR 双方をモデルに入力した場合の出力例を示す．正解要約，スライド画像は下記[5)] より抜粋した．特筆すべき例として，ASR/OCR 単体のみを入力した場合では生成されなかった語が要約に含まれた．スライドとトークの両方で言及されることで，より重要な語として認識された結果と考える．一方で，いずれかの入力には記載があるものの ASR と OCR の双方を入力しても生成されない内容も存在した．

PresenSum の課題と展望　本研究で得られた知見により，音声・画像から得られるテキスト情報を用いた要約について有望な結果が得られたものの，レイアウト・画像情報の利用については，fine-tune 時に特徴量を追加するだけでは十分な効果が得られないことがわかった．スライドの意味領域（タイトル・図表など）の構造的な理解や，事前学習などによる追加特徴量の理解の強化などの検討が必要である．また，音声認識誤りの影響の軽減や音声の重要箇所特定などのため，トーク音声のテキスト以外の特徴量の効果についても検証の余地がある [25].

5　関連研究

プレゼンテーションに関連する要約研究として TalkSum [8]，VT-SSum [9]，AVIATE [10] などがある．[8, 9] はトーク音声の ASR テキストのみを入力し，抽出型の要約タスクを行っている．[9] は抽出型要約の正解データ作成にスライド画像の OCR を利用した．AVIATE [10] はプレゼンテーションの動画から ASR や OCR などを行い，生成型のテキスト要約を行っている．しかし，OCR については動画中の粗い画像情報から抽出した上限 500 単語という部分的な利用にとどまっており，高精細なスライド画像の OCR テキスト全体やテキストのレイアウト等の詳細情報を利用を想定する我々の設定とは異なる．

6　おわりに

プレゼンテーションのトーク音声とスライド画像情報を用いて要約を行う PresenSum タスク提案し，データセットの収集，分析及びモデル評価を行った．PresenSum が従来の要約タスクと異なる特徴をもつこと，トーク音声とスライド画像のテキストが補完関係にあることを示し，要約モデルの入力として双方を用いることの有効性を確認した．また，OCR テキストのレイアウト・画像情報の効果についても初めて検証を行った．PresenSum は言語・音声・画像の融合的な理解が必要となるタスクであり，ビデオ会議の要約など幅広い分野に応用できる．

5）https://videolectures.net/icml09_frank_mac/

参考文献

[1] Hui Lin and Vincent Ng. Abstractive summarization: A survey of the state of the art. In **AAAI**, pp. 9815–9822, 2019.

[2] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. **J. Mach. Learn. Res.**, Vol. 21, pp. 140:1–140:67, 2020.

[3] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In **ACL**, pp. 7871–7880, 2020.

[4] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. PEGASUS: pre-training with extracted gap-sentences for abstractive summarization. In **ICML**, 2020.

[5] Anubhav Jangra, Adam Jatowt, Sriparna Saha, and Mohammad Hasanuzzaman. A survey on multi-modal summarization. **arXiv preprint arXiv:2109.05199**, 2021.

[6] Ramon Sanabria, Ozan Caglayan, Shruti Palaskar, Desmond Elliott, Loïc Barrault, Lucia Specia, and Florian Metze. How2: A large-scale dataset for multimodal language understanding. In **NeurIPS@ViGIL**, 2018.

[7] Zhou Y. Zhang J. Li H. Zong C. Li C. Zhu, J. Multimodal summarization with guidance of multimodal reference. In **AAAI**, pp. 9749–9756, 2020.

[8] Guy Lev, Michal Shmueli-Scheuer, Jonathan Herzig, Achiya Jerbi, and David Konopnicki. Talksumm: A dataset and scalable annotation method for scientific paper summarization based on conference talks. In **ACL**, pp. 2125–2131, 2019.

[9] Tengchao Lv, Lei Cui, Momcilo Vasilijevic, and Furu Wei. Vt-ssum: A benchmark dataset for video transcript segmentation and summarization. **arXiv preprint arXiv:2106.05606**, 2021.

[10] Yash Kumar Atri, Shraman Pramanick, Vikram Goyal, and Tanmoy Chakraborty. See, hear, read: Leveraging multimodality with guided attention for abstractive text summarization. **Knowl. Based Syst.**, Vol. 227, , 2021.

[11] Minesh Mathew, Dimosthenis Karatzas, and CV Jawahar. Docvqa: A dataset for vqa on document images. In **WACV**, pp. 2200–2209, 2021.

[12] Ryota Tanaka, Kyosuke Nishida, and Sen Yoshida. Visualmrc: Machine reading comprehension on document images. In **AAAI**, pp. 13878–13888, 2021.

[13] Minesh Mathew, Viraj Bagal, Rubèn Tito, Dimosthenis Karatzas, Ernest Valveny, and CV Jawahar. Infographicvqa. In **WACV**, pp. 1697–1706, 2022.

[14] Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In **NIPS**, pp. 1693–1701, 2015.

[15] Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In **EMNLP**, pp. 1797–1807, 2018.

[16] Arman Cohan, Franck Dernoncourt, Doo Soon Kim, Trung Bui, Seokhwan Kim, Walter Chang, and Nazli Goharian. A discourse-aware attention model for abstractive summarization of long documents. In **NAACL**, 2018.

[17] Max Grusky, Mor Naaman, and Yoav Artzi. Newsroom: A dataset of 1.3 million summaries with diverse extractive strategies. In **ACL**, pp. 708–719, 2018.

[18] Yang Xu, Yiheng Xu, Tengchao Lv, Lei Cui, Furu Wei, Guoxin Wang, Yijuan Lu, Dinei A. F. Florêncio, Cha Zhang, Wanxiang Che, Min Zhang, and Lidong Zhou. Layoutlmv2: Multi-modal pre-training for visually-rich document understanding. In **ACL/IJCNLP**, pp. 2579–2591, 2021.

[19] TY Lin, M Maire, S Belongie, J Hays, P Perona, D Ramanan, P Dollar, and CL Zitnick. Microsoft coco: Common objects in context. In **ECCV**, pp. 740–755, 2014.

[20] Ren, Shaoqing, He, Kaiming, Girshicka, Ross, Sun, and Jian. Faster r-cnn: Towards real-time object detection with region proposal networks. In **NIPS**, pp. 91–99, 2015.

[21] Lei Jimmy Ba, Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. **arXiv preprint arXiv:1607.06450**, 2016.

[22] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In **Text summarization branches out@ACL**, pp. 74–81, 2004.

[23] Iz Beltagy, Matthew E. Peters, and Arman Cohan. Longformer: The long-document transformer. **arXiv preprint arXiv:2004.05150**, 2020.

[24] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In **ICLR**, 2019.

[25] Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. wav2vec 2.0: A framework for self-supervised learning of speech representations. In **NeurIPS**, 2020.

A　補足資料

ASR 単体，OCR 単体，ASR と OCR の双方を入力した場合の要約出力例を図 5 に示す（図 4 は ASR+OCR と参照要約の抜粋である）．正解要約テキストとスライド画像は下記[6] より抜粋した．ASR 単体は意味が不明瞭な箇所があるが，ASR と OCR を合わせて入力することで改善がみられる．

OCR only

We present a multi- Assignment Clustering approach for Boolean data. Our approach is based on a probabilistic representation of cluster centroids, which are characterized by disjunctions of Booleanemissions. We compare the accuracy of different methods at different noise levels, and compare them with other multi-assignment methods. We show that our approach outperforms the state-of-the-art methods in terms of accuracy. We also compare the performance of different multi- assignment methods with other methods. Our results show that the multi-Assignments approach is more accurate than the state of the art methods, and that it is more robust to noise than the other approaches. Furthermore, our approach is able to estimate cluster centroid mismatch at different levels of noise.

ASR only

In this paper, we present a probabilistic model for clustering Boolean data where similar objects are allowed to belong to more than one cluster at the same time. The model is based on the notion that the objects and the permissions are independent from each other. We show that the model can be used to infer the permissions of a given object by inferring the permissions from its input data. We demonstrate that the proposed model outperforms the state-of-the-art clustering methods in terms of accuracy and noise. We also demonstrate the effectiveness of our model on a real-world dataset from a Swiss bank. Our experiments on synthetic data demonstrate that our proposed model can outperform the state of the art clustering techniques. We further demonstrate the applicability of our method on real world datasets.

OCR + ASR

In this paper, we present multi- Assignment Clustering for Boolean Data, a method of pooling data where similar objects are allowed to belong to more than one cluster at the same time. We show that multi assignment clustering can be applied to a variety of applications, such as **role mining** and role-based **access control**. We demonstrate the effectiveness of our method on real-world datasets and compare it to several combinatorial methods on synthetic data. We also show that our method outperforms the state-of-the-art methods in terms of accuracy and noise reduction, and that it can also be used to detect wrong or erroneous assignments of objects in real world datasets. The results of our experiments on synthetic datasets show that the method is able to outperform the state of the art.

Reference

Conventional clustering methods typically assume that each data item belongs to a single cluster. This assumption does not hold in general. In order to overcome this limitation, we propose a generative method for clustering vectorial data, where each object can be assigned to multiple clusters. Using a deterministic annealing scheme, our method decomposes the observed data into the contributions of individual clusters and infers their parameters. Experiments on synthetic Boolean data show that our method achieves higher accuracy in the source parameter estimation and superior cluster stability compared to state-of-the-art approaches. We also apply our method to an important problem in computer security known as **role mining**. Experiments on real-world **access control** data show performance gains in generalization to new employees against other multi-assignment methods. In challenging situations with high noise levels, our approach maintains its good performance, while alternative state-of-the-art techniques lack robustness.

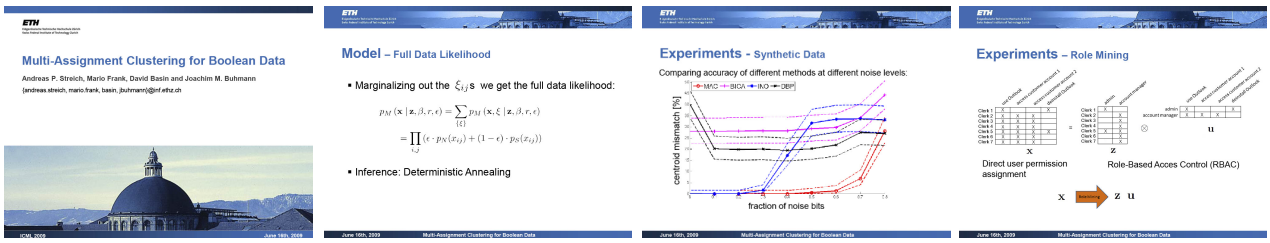

図 5　ASR 単体，OCR 単体，ASR と OCR の双方を入力した場合の要約出力例．赤文字が ASR や OCR 単体入力においても生成された内容．緑文字が ASR と OCR を両方入れた時のみ生成された内容．青文字はすべての要約で生成されなかった内容．

6）https://videolectures.net/icml09_frank_mac/