

自然言語理解の 社会実装に向けて

産業技術総合研究所

麻生英樹

総務省次世代人工知能社会実装WG

2017.2.14

自然言語理解

- 複雑な言語の使用は人の知能の最大の特徴
- 人間と同等なレベルで自然言語を理解し、操れるシステムの実現は Turing test 以来の人工知能研究のグランドチャレンジ
- 人間レベルの自然言語理解ができれば：
 - 大量に蓄積された言語的知識や文献情報の活用拡大
 - センサ情報の活用拡大、データからの知識獲得
 - ロボットとの対話的な意思疎通
 - 人間と相互理解できる人工知能、説明可能な人工知能の実現

自然言語理解研究の方向性

- 論理的な意味の理解
 - 同義文、言い換えの判定や生成
 - 含意関係認識
 - Semantic Parsing (テキスト → 意味表現)
 - 知識ベース検索
 - 知識を使った推論による問題解決
- 感覚・運動情報との結びつけ
 - シンボル・グラウンディング (記号接地)、記号創発
 - 画像・動画と文章の相互変換
 - 画像・動画に関する質問応答
 - 言語からの運動生成、など

含意関係認識

- 文A が文B を含意する（文A が正しいときには文B も正しい）か否かを判定

A：プロメテウスは人類に火を渡し、
磔にされた。

B：人間はプロメテウスから火を渡された。

含意関係認識

- 語彙的・構文的言い換えや論理的含意関係の判定の場合
- 一般的知識を使った推論も必要な場合

A：延喜・天曆の治とは、平安時代中期（10世紀）の醍醐・村上両天皇の治世を聖代視した呼称である。

B：10世紀に、延喜・天曆の治と後によばれる政治が行われた。

含意関係認識

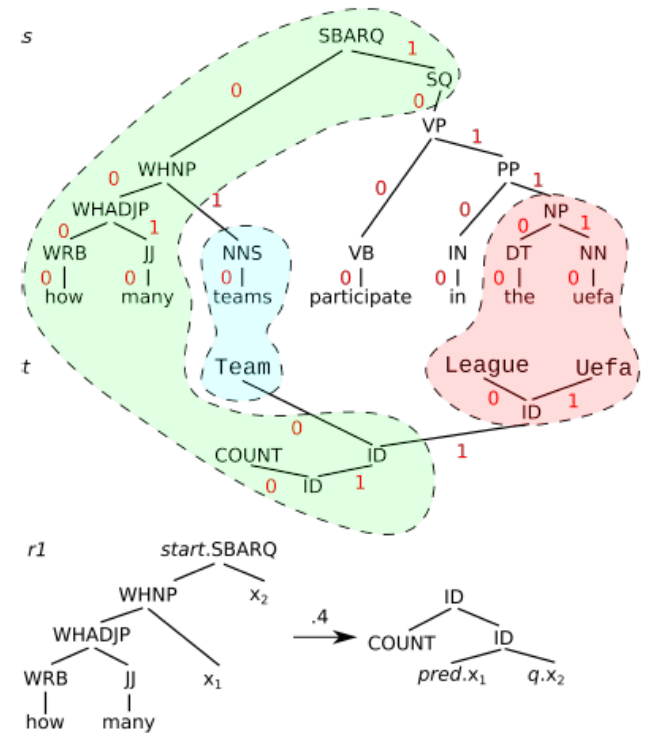
● NTCIR RITE-VAL タスク

- 大学入試センター試験の社会科学科目の正誤を問う設問を題材とした含意関係認識タスク
- 2014年開催の NTCIR11 の優勝チームの正解率は 80%弱
- 答えの導き方は文 A と文 B から抽出した評価特徴量の類似度に基づくもので、人間とはかなり異なる
- 論理的な意味表現や推論は使っていない
- 分散意味表現の導入の試みも行われている

石井、宮下：含意関係認識のための機械学習と全文検索, Unisys技報, 2016.

Semantic Parsing

- テキスト → データベース検索
や論理的推論に使えるフォー
マルな意味表現
- 事例からの変換ルールの学習
 - 教師あり学習：テキストと
意味表現ペアからの変換
ルール学習
 - 半教師あり学習：テキスト
とデータベース検索の回答
ペアからの変換ルール学習

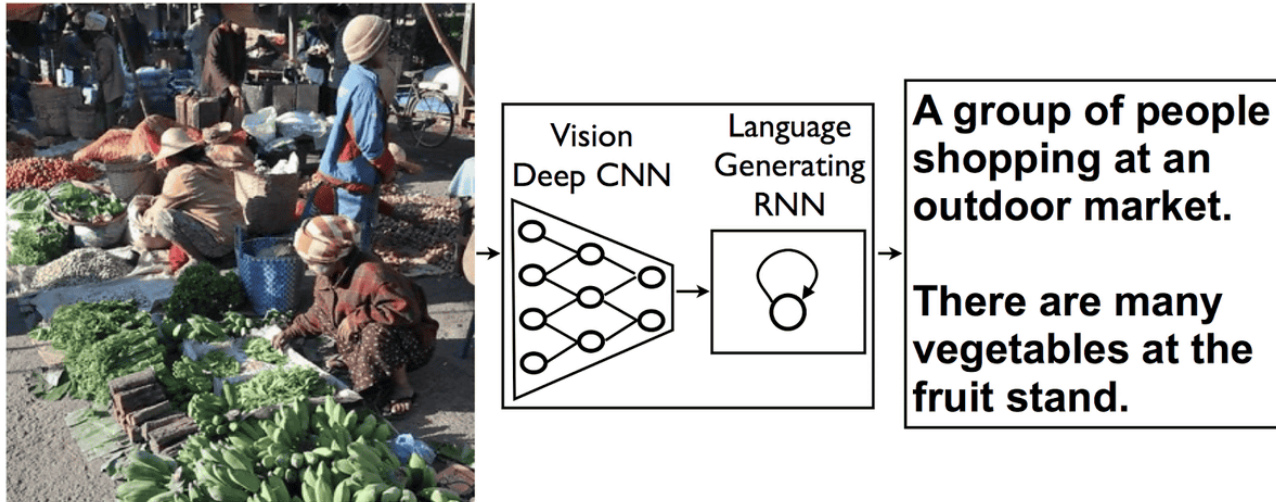


Martinez-Gomez, P., Miyao, Y.: Rule extraction for tree-to-tree transducers by cost minimization, EMNLP 2016.

感覚・運動情報との結びつけ

- 深層学習および分散意味表現の発達により研究が急速に進展中
- 精度はまだまだ人間には及ばない
- タスクの例
 - 画像・動画と文章の相互変換
 - 画像・動画に関する質問応答
 - 言語からの運動生成、など

画像からの説明文生成



Vinyals+ : Show and tell: a neural image caption generator, 2014.



A woman is throwing a frisbee in a park.

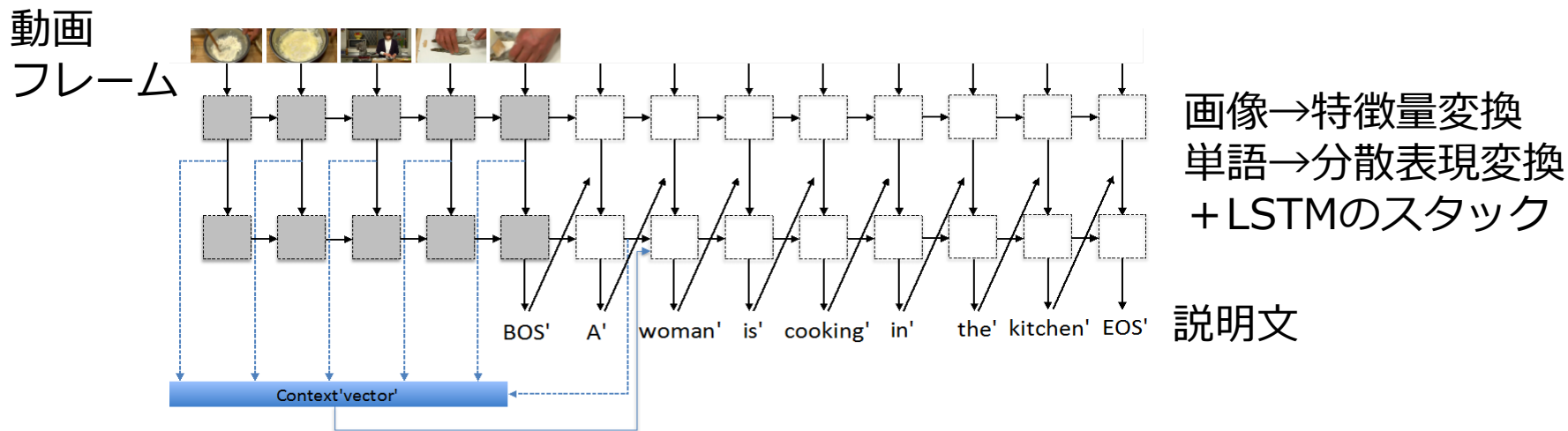
A dog is standing on a hardwood floor.

A stop sign is on a road with a mountain in the background.

Xu+ : Show, attend and tell: neural image caption generation with visual attention, 2015.

動画からの説明文生成

- 動画への説明文付与：
Seq2Seq モデルへの時間的注意の導入



Laokulrat, N. et al.: Generating video description using sequence-to-sequence model with temporal attention, COLING 2016.

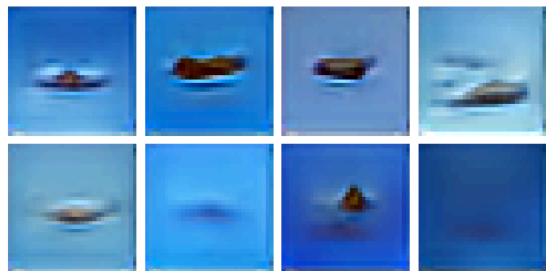
動画からの説明文生成

- Microsoft Video Description dataset
1,970 short clips from YouTube BLUE=0.437
- Montreal Video Annotation dataset
46,589 short clips from 92 movies BLUE=0.008

動画の出典：YouTube



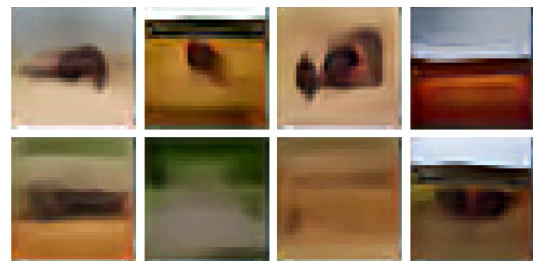
文章からの画像生成



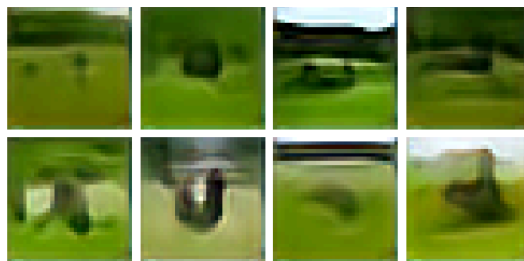
A very large commercial plane flying in blue skies.



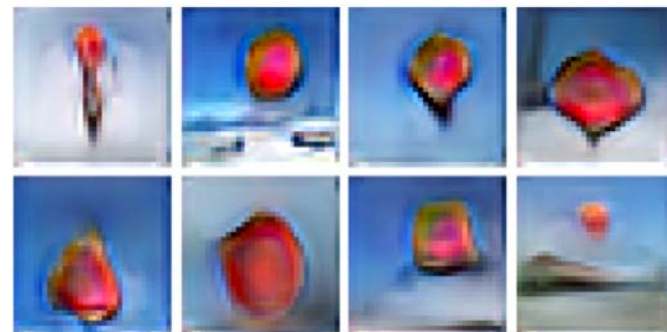
A very large commercial plane flying in rainy skies.



A herd of elephants walking across a dry grass field.



A herd of elephants walking across a green grass field.

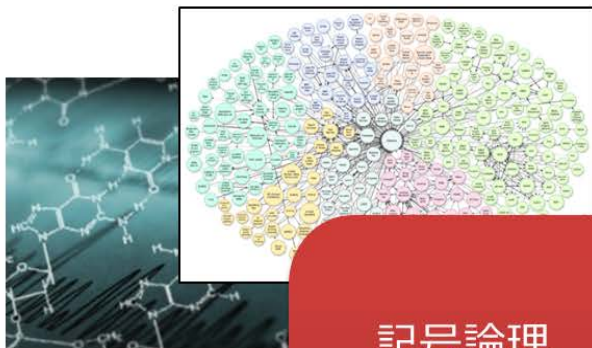


A stop sign flying in blue skies.

Mansimov E. et. al: Generating images from captions with attention, NIPS 2015 Reasoning, Attention, Memory (RAM) Workshop, 2015.

データ・知識融合

知識ベース・オントロジー



センサーデータ・行動データ



意味表現

記号論理

```
SELECT DISTINCT ?uri
WHERE {
  ?uri rdf:type
    yago:StatesOfGermany .
  ?uri dbp:rulingParty
    res:SDP_of_Germany .
}
```

分散表現



世界モデル



自然言語テキスト



画像・映像

自然言語処理・画像処理基盤ツール

図版提供：宮尾祐介招聘研究員 (AIRC/NII)

自然言語理解の社会実装の困難さ

- 浅い理解（たとえば単語の出現頻度）でもそれなりに役立つ
- 理解度を向上させるコストに結果のベネフィットが見合わない
 - 分野ごとの辞書構築などに大量の人手が必要
 - 大量のラベルつき学習用データが必要
 - 「深い理解」を前面に出すと人間の要求水準は高い
- 映画やテレビドラマを見せておくだけで勝手に学習してくれるとありがたいが・・・

予測に基づく学習

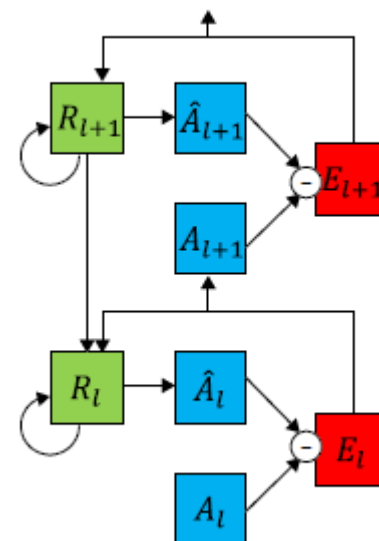
- 映画の中の 1 秒後の行動を予測

	Forecast 1 Second Before Action Starts	Ground Truth Action (not observed)
Successful Predictions	<p>Kiss</p>	<p>Kiss</p>
	<p>High Five</p>	<p>High Five</p>
	<p>Hug</p>	<p>Hug</p>

Vondrick, C. et al.: Anticipating the future by watching unlabeled video, CVPR 2016.

予測に基づく学習

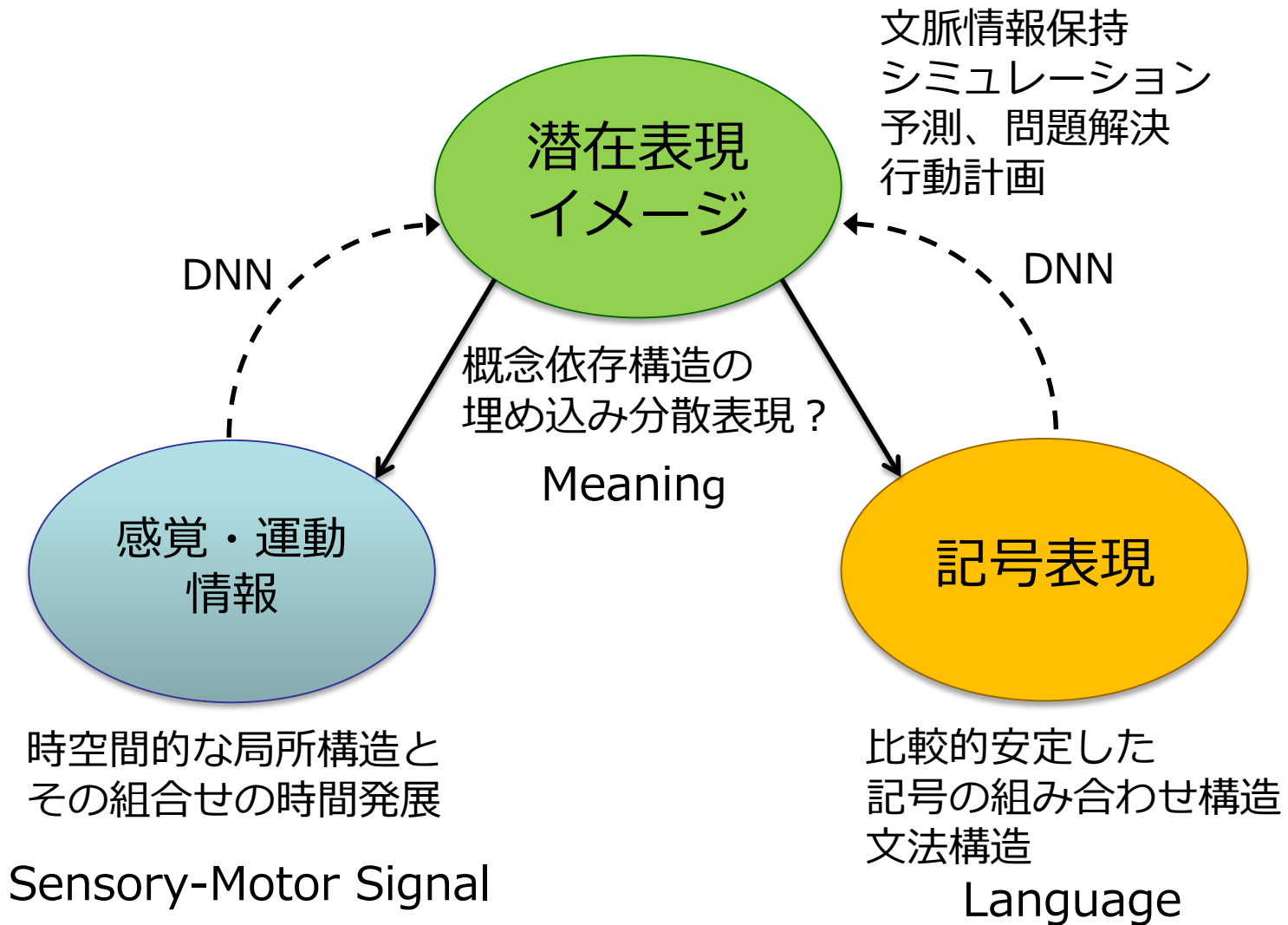
- 数フレーム先の映像を生成



Lotter, W. et al.: Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning, arXiv 2016.

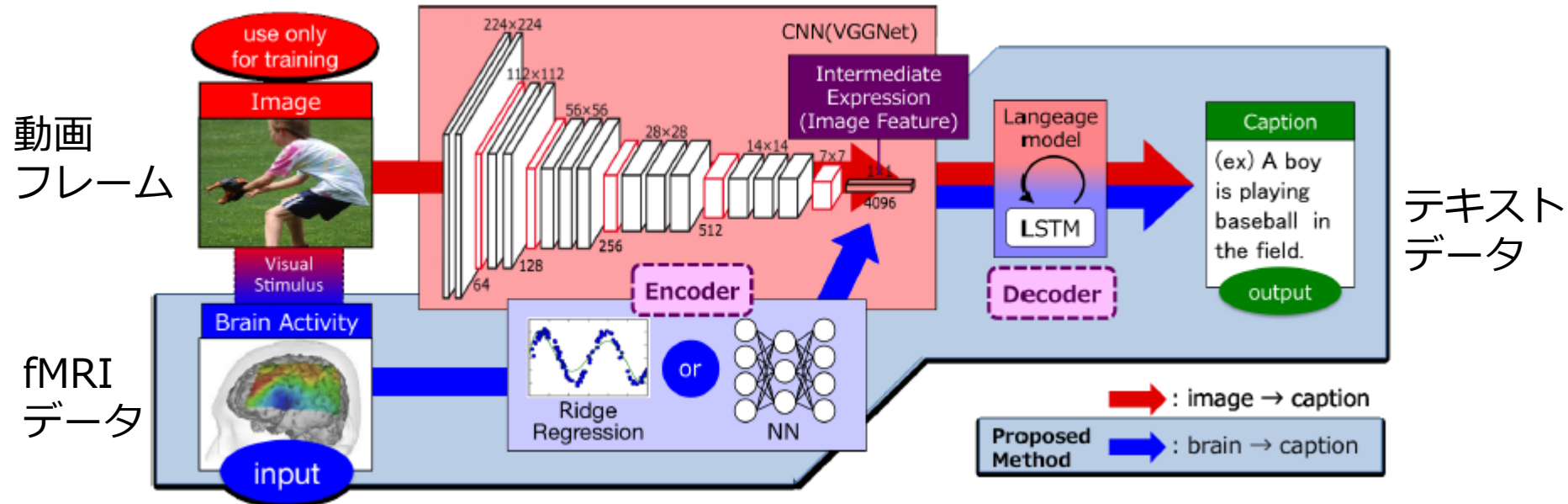
- 予測学習で得られる潜在表現を言語表現と結び付けられないか？

潜在表現の学習と利用



脳情報通信研究との連携

- 脳内情報表現と言語表現の関連づけ



Matsuo, E. et al.: Generating natural language descriptions for semantic representations of human brain activity, ACL 2016 Student Workshop, 2016.

まとめ

- 意味理解技術の技術的課題と研究開発状況
 - 論理的意味の理解
 - 感覚・運動情報との結びつけ
 - 脳情報通信研究との連携
 - 社会実装に向けて、様々な研究が進展中
- 次世代人工知能技術の社会実装に向けて
 - 短期：現在の技術を改良して有用なタスクに適用
 - 長期：次のブレークスルー、汎用性が高い技術の開拓
 - 両者をうまく組み合わせる枠組み？