

次世代メディア処理AI技術

人間は、1つの脳に五感で得られた知識を蓄積することができ、その知識を共通的に参照しながら様々な情報を処理することが可能です。人間はこのような情報処理機構を持つからこそ、効率的に知識を蓄えて成長することができ、その知識を活用して世の中の様々な情報を正確に理解することができます。NTTでは、AIにも人間と同様の情報処理機構を持たせることで、人間のような効率的な「学習」や、マルチメディアの同時処理による高精度かつ総合的な「推論」が可能になると期待して、MediaGnosisの研究開発を進めています。

MediaGnosisでは、音声音響処理・画像映像処理・自然言語処理といったマルチメディアの情報処理を統合的にオールインワンで扱うことで、これまでよりも効率的な「学習」と、高精度かつ総合的な「推論」を実現します。

MediaGnosisの特徴は、「知識統合型アーキテクチャ」により、人間のように1つの脳(AIにおいてはモデル)の中で、各種メディア処理AI機能を統合的にサポートするモデリングを行うことができる点です。具体的に「知識統合型アーキテクチャ」では、図に示すように、入力モーダル(音声音響情報・画像映像情報・自然言語情報、など)、および出力対象(数値・ベクトル・ラベル・テキスト、など)の両者に対して、複数の機能間で同様の役割を持つ情報処理機構(理解部)を、複数の機能間で共通化します。これにより、1つのモデルの中に様々なAI処理機能を統合することが可能となります。

本技術により、①これまでよりも少ないデータ量で、人のように効率的な知識獲得が可能な「学習」が実現できます。例えば、音声認識の学習を行うことで、画像認識や機械翻訳の性能改善につなげることが可能となります。そして、②複数の機能を同時に駆動した「推論」が実現できます。例えば、声質とその言葉の内容、表情を同時に考慮して人の感情を推論することが可能となります。

●MediaGnosisの知識統合型アーキテクチャの概略図



MediaGnosis