

翻訳バンクで ニューラル機械翻訳を 多分野化する

NICT

隅田 英一郎

Big Dataから深層学習で構築する ニューラル機械翻訳は高精度

Big Data

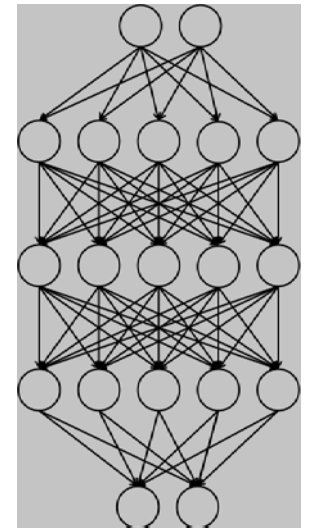
人類は、国家や地域、民族や世代など、あらゆる境界を越えて、相互の理解を深め、知恵を交わすなかで、発展してきました。コミュニケーションは人類社会を支えるもっとも重要な活動であり、情報通信技術はそのコミュニケーションを支える基礎であります。情報通信技術はまた、人類の高度な知的活動と経済活動を支える基盤でもあります。

Humanity has achieved progress as it has deepened its mutual understanding and shared its wisdom, overcoming barriers due to national, regional, ethnic, generational, and other differences. Communication is the most critical activity in human society, and information and communications technology (ICT) is the basis of that communication. ICT is also the infrastructure that supports humanity's advanced intellectual and economic activities.

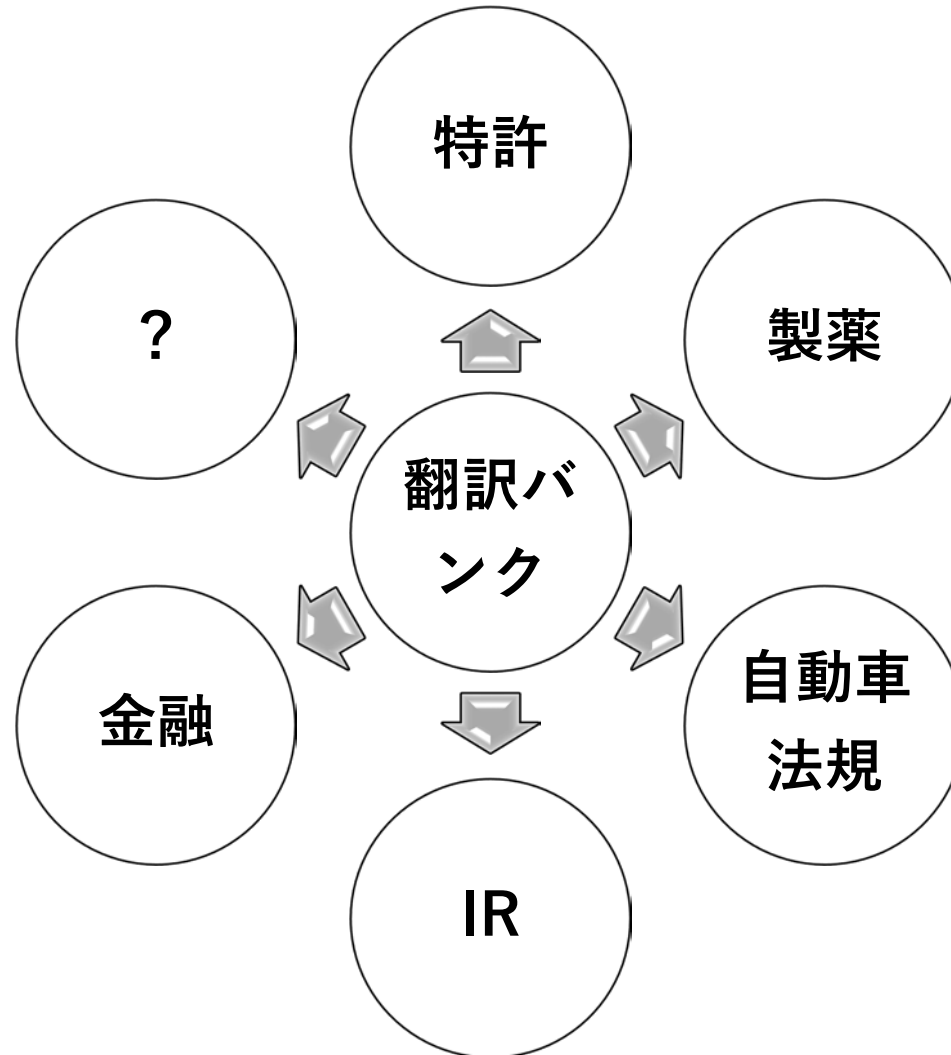
<http://www.nict.go.jp/およびhttp://www.nict.go.jp/en/index.html>

深層学習

ニューラルネット



多分野化が加速



日興証券と金融特化型を開発

BLEU スコア	解釈
> 60	人が翻訳した場合よりも高品質であることが多い
50~60	非常に高品質で、適切かつ流暢な翻訳
40~50	高品質な翻訳
30~40	理解できる、適度な品質の翻訳
20~29	主旨は明白であるが、文法上の重大なエラーがある
10~19	主旨を理解するのが困難である
< 10	ほとんど役に立たない

13.1 点から
42.8 点に
改善！

入力	H形鋼と熱延鋼板に関しては、鉄屑市況の上昇分をカバーする水準まで上昇しそうだ。
汎用訳	The prices of H-shaped steel and hot-rolled steel are expected to rise to levels that will cover the rise in the steel scrap market.
金融特化訳	For H-beams and hot-rolled sheet, we expect prices to rise to levels that offset the rise in steel scrap prices.