Challenges in Adaptive Neural Machine Translation

Marcello Federico
MMT Srl / FBK Trento, Italy

MODERN MT

FONDAZIONE BRUNO KESSLER

Our Adventures with ModernMT (2015-2017)

MODERN MT

FONDAZIONE BRUNO KESSLER

Symbiotic Human and Machine Translation

MT seamlessly
- adapts to user data
- learns from post-editing

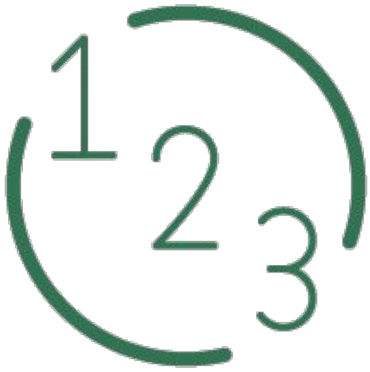
user enjoys
- enhanced productivity
- better user experience

Usable technology for the translation industry

- **easy** to install and deploy
- **fast** to set-up for a new project
- **effective**, also on small projects
- **scalable** with data and users
- works with commodity hardware

The Modern MT way

(1) connect your CAT with a **plug-in**
(2) drag & drop your **private** TMs
(3) start translating!



MODERN MT

FONDAZIONE BRUNO KESSLER

5

Modern MT in a nutshell

zero training time
adapts to context
learns from user corrections
scales with data and users

MODERN MT

FONDAZIONE BRUNO KESSLER

6

Fast training

Training data is a **dynamic** collection of Translation Memories

At any time:

- new TMs are **added**
- existing TMs are **extended**

Training time comparable to uploading time!

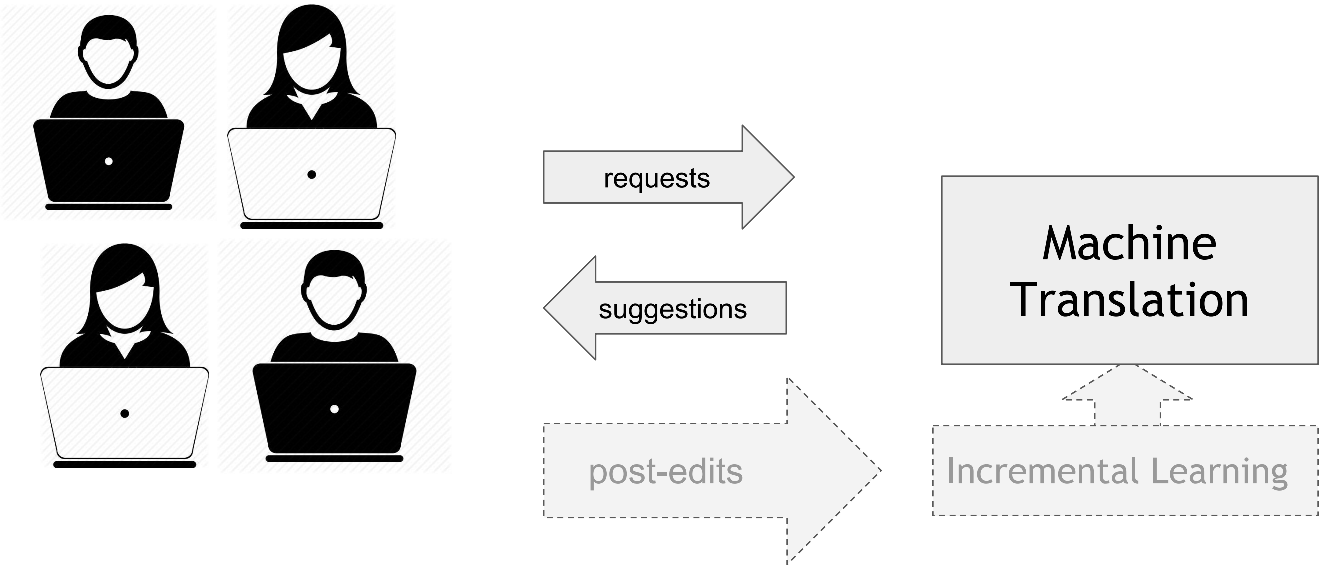
Context aware translation

SENTENCE

party

CONTEXT	CONTEXT
We are going out.	We approved the law
TRANSLATION	**TRANSLATION**
fête	parti

Incremental learning



Core technology [original plan]

context analyser
phrase-based decoder
adaptive models
incremental structures
parallel processing

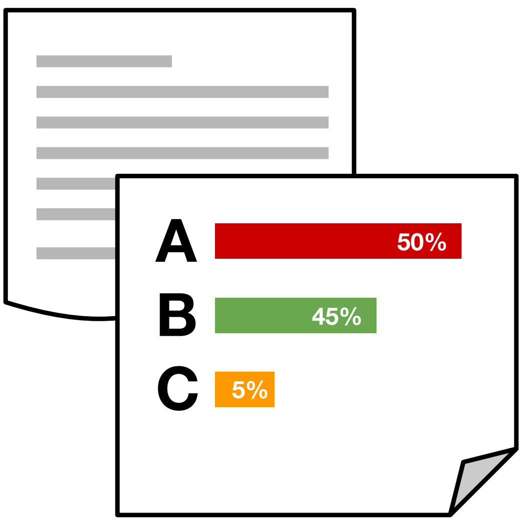
Language support

- **45** languages
- **fast pre-/post-processing**
- **simple interfaces**
- **tags** and **XML** management
- localization of **expressions**
- TM **cleaning**

MODERN MT

Simple. Adaptive. Neural.

Context Analyzer

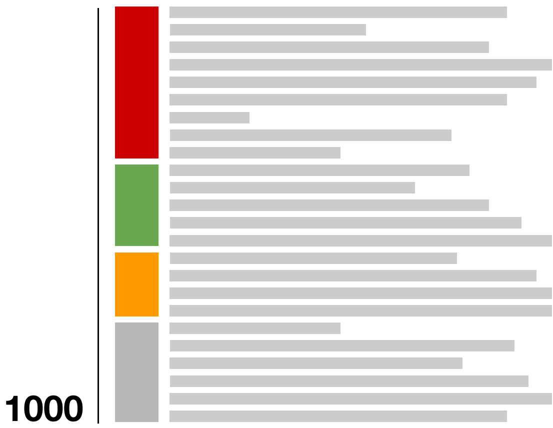


- analyze input text
- retrieve best matching TMs
- compute matching scores
- **dynamic structure**

MODERN MT

Simple. Adaptive. Neural.

Adaptive Phrase Table

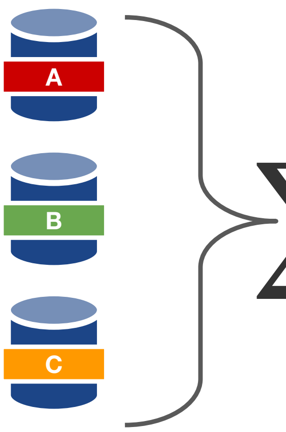


Suffix Array with
Ranked Sampling

- suffix array indexed with TMs
- phrases sampled on demand
- priority sampling over TMs
- **dynamic structure**

MODERN MT — Simple. Adaptive. Neural.

Adaptive Language Model

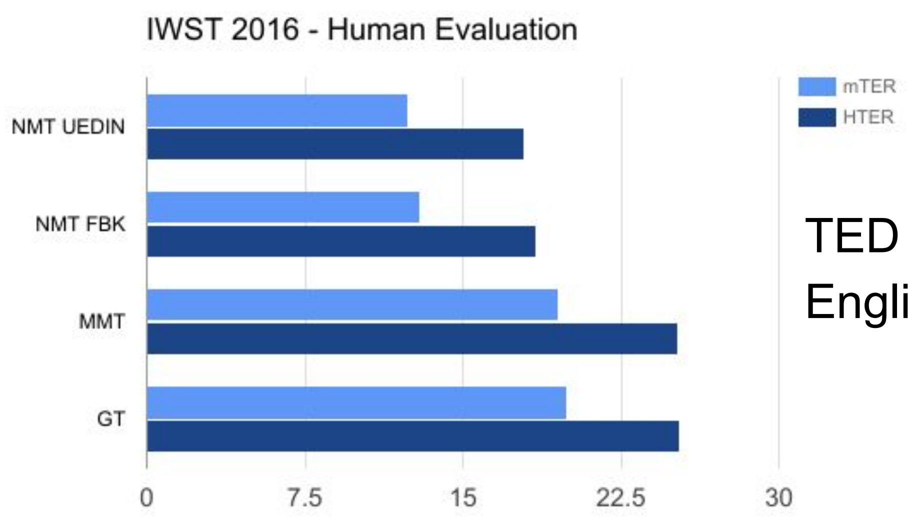


$$\sum w \cdot p$$

- large static background model
- n-grams stats indexed with TMs
- combination of *active* TM LMs
- TM LMs computed on the fly
- **dynamic structure**

MODERN MT — Simple. Adaptive. Neural.

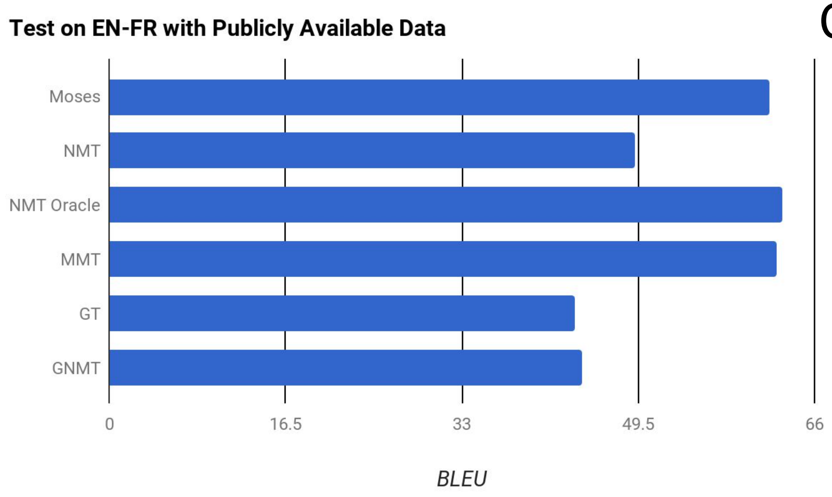
Statistical vs. Neural MT



TED Talks
English-French

M. Cettolo, et al. (2016), *The IWSLT 2016 Evaluation Campaign*, IWSLT.

Second Prototype (0.14 January 2017)



Open benchmark:

- Training speed:
  12x Moses - 100x NMT
- MT quality (BLEU):
  +1 vs Moses
  -0.5 vs NMT Ada

Domains: ECB, Gnome, JRC, KDE, OpenOffice, PHP, Ubuntu, UN-TM

What happened

Research on adaptive neural MT

Believed PBMT was competitive on technical translation

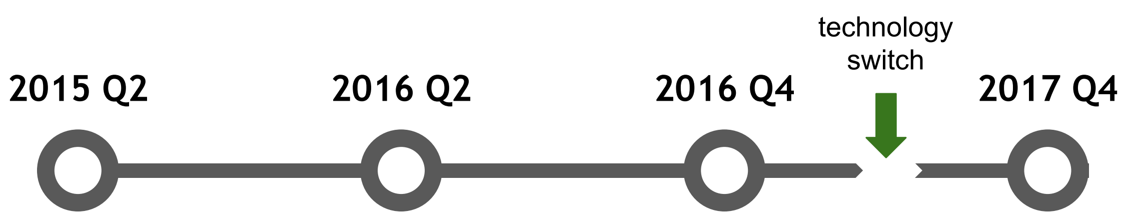
Finally realised superiority of NMT quality

Completed PBMT release and **switched to NMT**

Data collection for 14 translation directions

MODERN MT Simple. Adaptive. Neural.

Roadmap from last review meeting

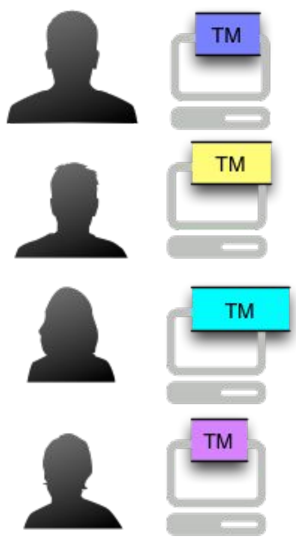


2015 Q2	2016 Q2	2016 Q4	2017 Q4
minimum viable product	first **alpha release**	first **beta release**	**final release**
context aware 1 lang pair	**fast training, context aware**, distributed, 1 lang pair	**online learning plug-in,** 3 lang pairs	**neural MT**, enterprise ready, 14 lang pairs

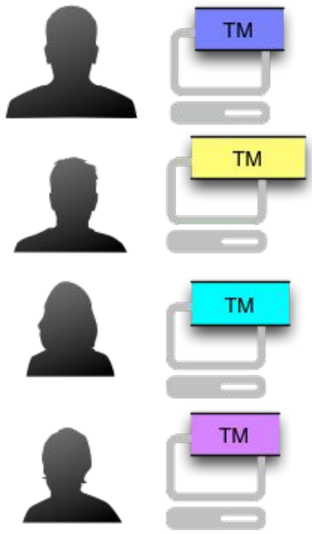
MODERN MT Simple. Adaptive. Neural.

Multi-Domain Neural MT

Multi-user scenario



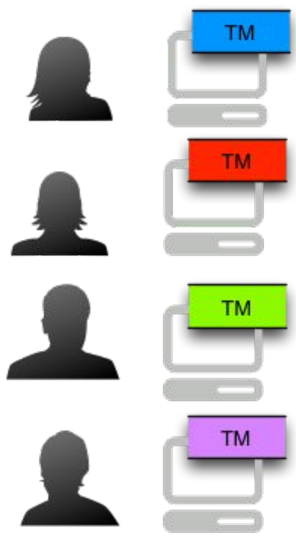
Multi-user scenario



MODERN MT

FONDAZIONE BRUNO KESSLER

Multi-user scenario



MODERN MT

FONDAZIONE BRUNO KESSLER

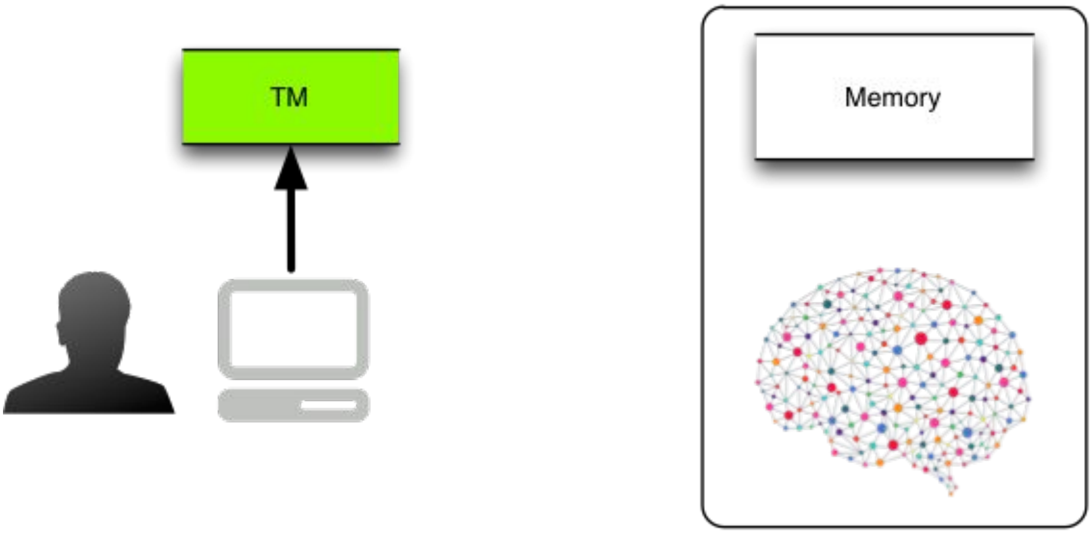
Adaptive Neural MT
(Adaptation *a priori*)

MODERN MT

FONDAZIONE BRUNO KESSLER

23

All we need is a memory

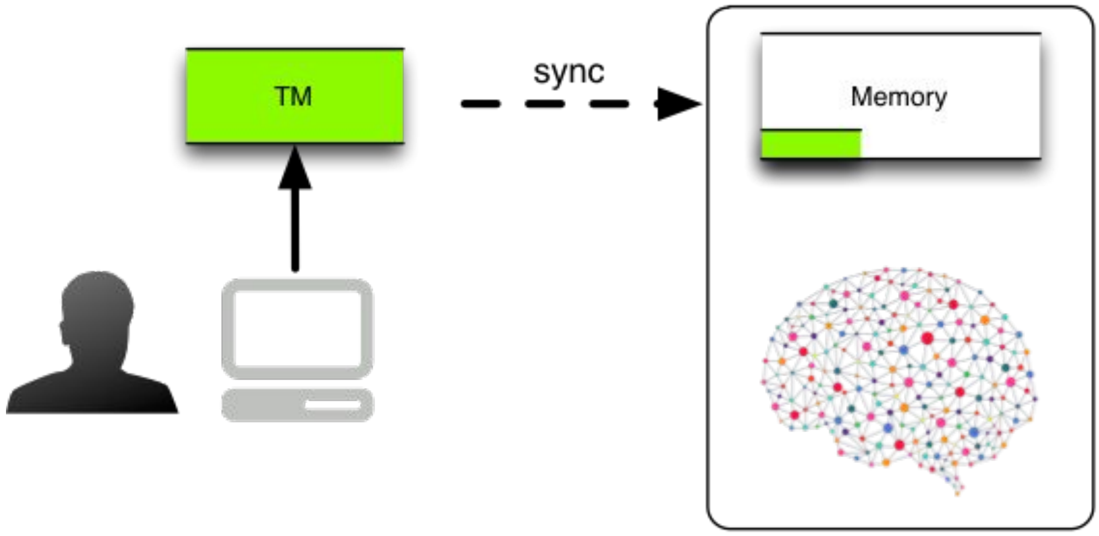


MODERN MT

FONDAZIONE BRUNO KESSLER

24

All we need is a memory

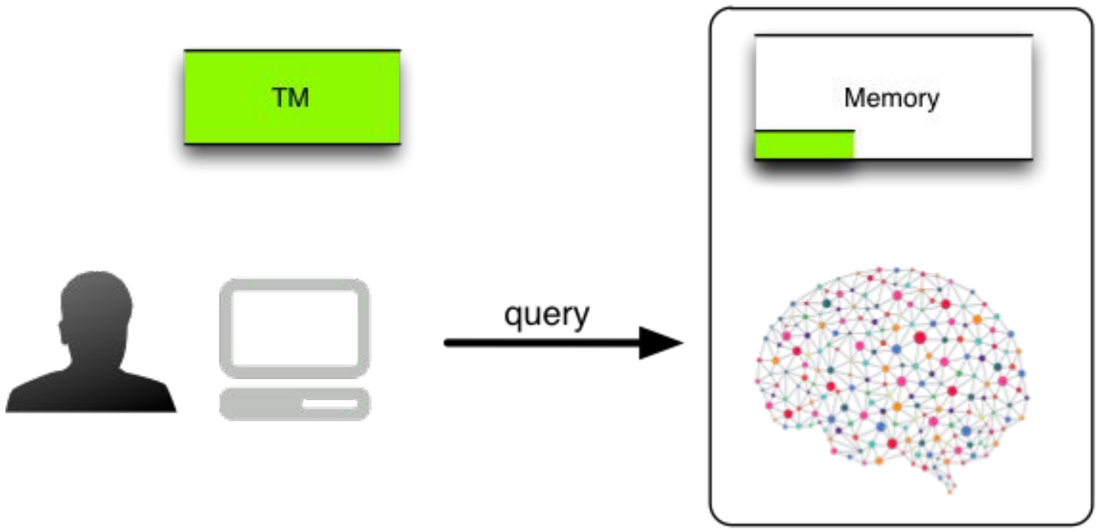


MODERN MT

FONDAZIONE BRUNO KESSLER

25

All we need is a memory



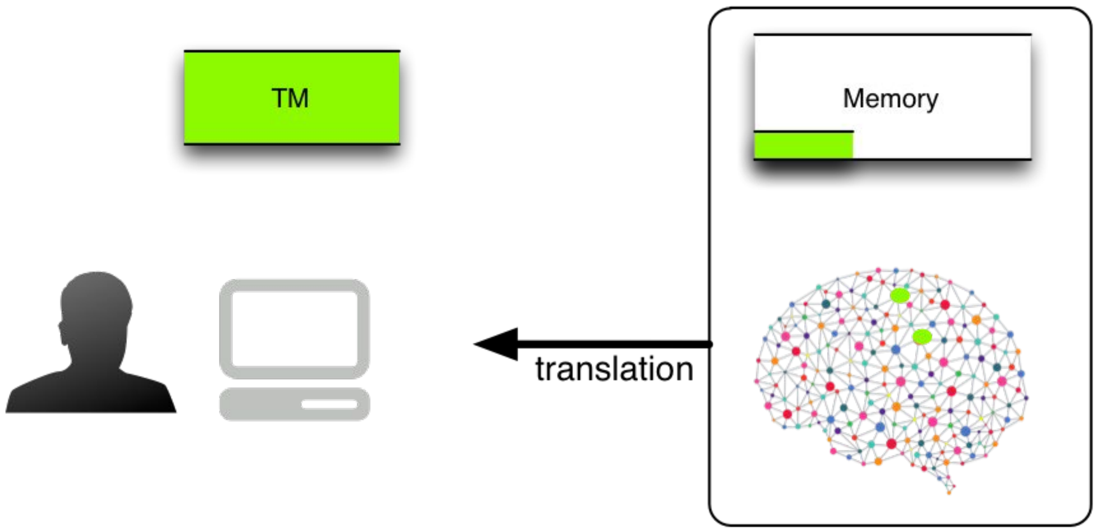
MODERN MT

FONDAZIONE BRUNO KESSLER

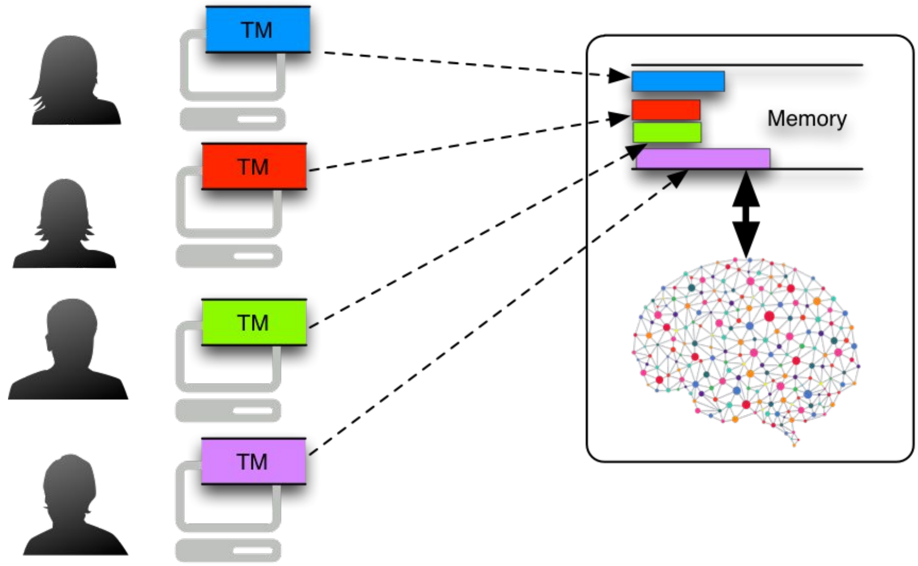
26

All we need is a memory

All we need is a memory



Multi-user adaptive NMT

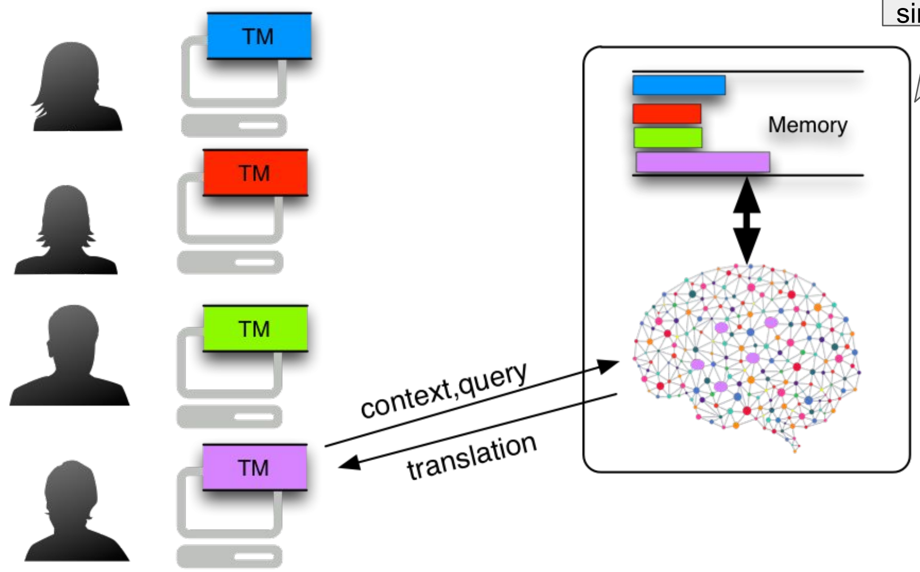


MODERN MT

FONDAZIONE BRUNO KESSLER

Multi-user adaptive NMT

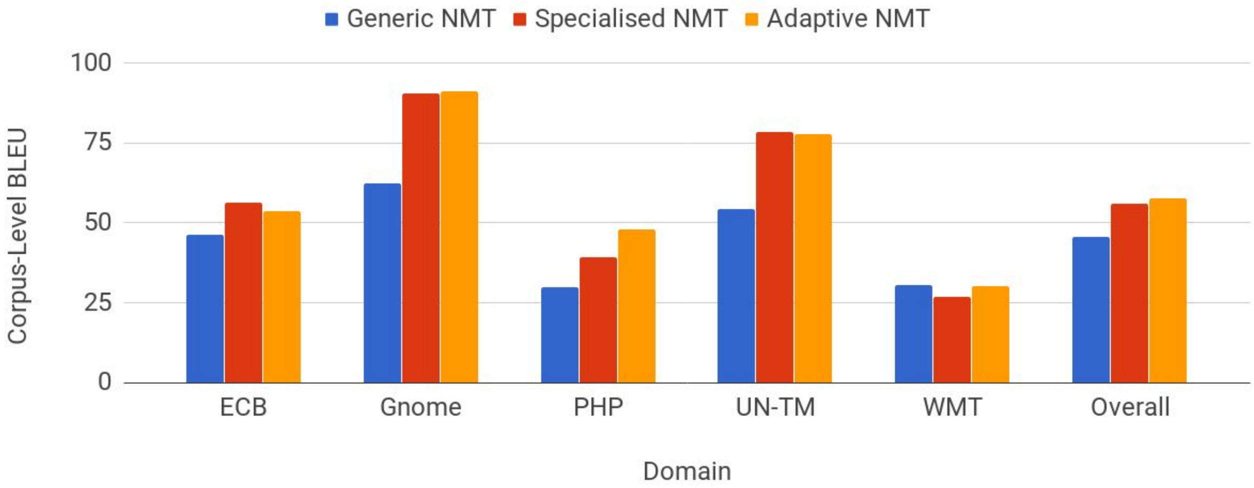
Instances are selected by combining context scores and similarity scores



MODERN MT

FONDAZIONE BRUNO KESSLER

Adaptation, too!



Farajian et al. (2017) “Multi-domain NMT through unsupervised adaptation”, *WMT*.

Production Systems

Timeline 2017

Sep: integration of MateCat
Oct: NMT code released
Nov: co-development
release of 14 engines
Dec: performance boost

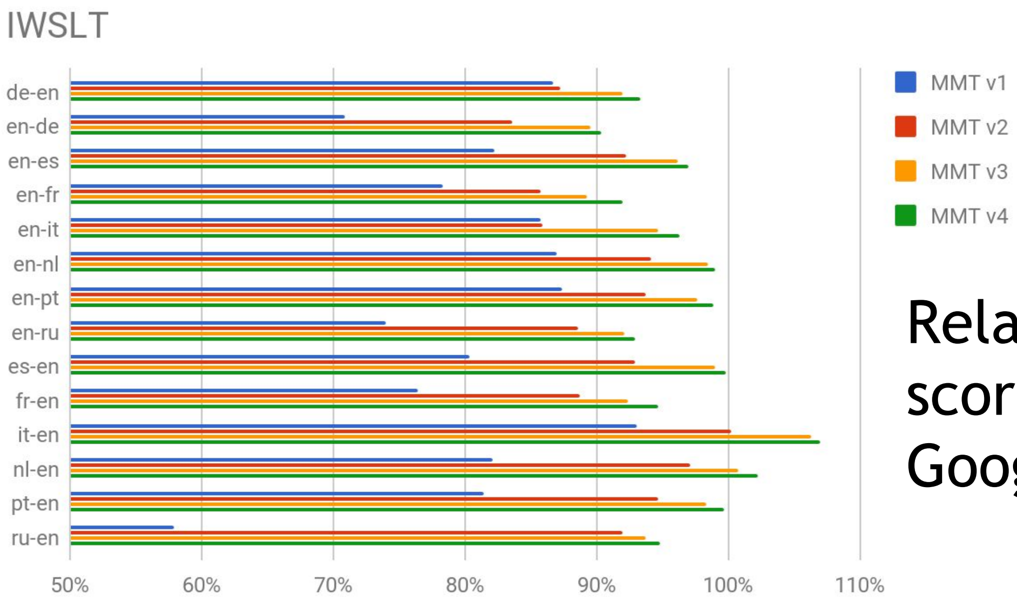
Marcello Federico
@marcfede

MMT team just released adaptive NMT in 14 directions for the @MateCat plugin!

5:10 PM - 30 Nov 2017

10 Retweets 22 Likes

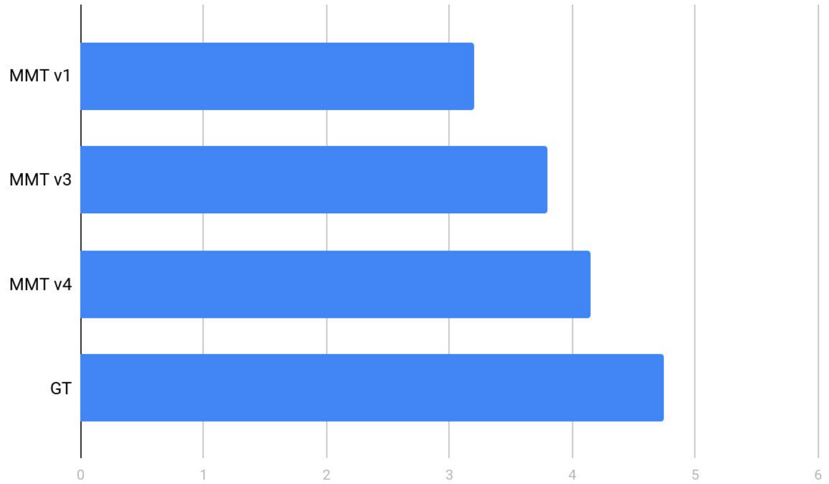
Automatic Evaluations



Relative BLEU scores wrt Google Translate

Micro HE Assessment

Progression in one month on English-Italian



Performance of generic MMT
1-6 scale
(w/o adaptation)

Quality Estimation

Quality Evaluation

MMT Eval 27/11/17 EN-IT index

Nonostante fosse ancora largamente sconociuto, Robert Redford fece il suo debutto sullo schermo in War Hunt (1962), affiancando John Saxon in un film ambientato durante gli ultimi giorni della Guerra di Corea.

OUTPUT **1**: = = Storia =

1 2 3 4 5 6

Add comment

OUTPUT **2**: Nonostante ancora una sconosciuta, Robert Redford fece il suo debutto dello schermo in War Hunt (1962), co-protagonista con John Saxon in un film organizzato durante gli ultimi giorni della guerra coreana.

1 2 3 4 5 6

Add comment

OUTPUT **3**: Mentre era ancora in gran parte sconosciuto, Robert Redford fece il suo debutto sul grande schermo in War Hunt (1962), recitando insieme a John Saxon in un set cinematografico durante gli ultimi giorni della Guerra di Corea.

1 2 3 4 5 6

Add comment

MODERN MT

FONDAZIONE BRUNO KESSLER

Quality Evaluation

MMT Eval 27/11/17 EN-IT index

L'Agenzia per il Rilevamento Geologico degli Stati Uniti (USGS) ha individuato l'epicentro del terremoto a 12.8 miglia (20.6 chilometri) di profondità, a circa 150 miglia (240 chilometri) da Bengkulu, Sumatra.

OUTPUT **1**: = = Note = = = = Bibliografia = = = = Altri progetti = = = = Collegamenti esterni = = * Sito ufficiale

1 2 3 4 5 6

Add comment

OUTPUT **2**: = = Note = = = = Altri progetti = = = = Collegamenti esterni = = * Sito ufficiale

1 2 3 4 5 6

Add comment

OUTPUT **3**: Lo United States Geological Survey (USGS) ha riportato l'epicentro del terremoto a 20,8 chilometri di profondità ea circa 150 miglia (240 chilometri) da Bengkulu, Sumatra.

1 2 3 4 5 6

Add comment

MODERN MT

FONDAZIONE BRUNO KESSLER

Noisy training data

EN: What history teaches us

IT: === Storia ==================================

MODERN MT   Simple. Adaptive. Neural.

Data Cleaning

We added a simple QE module to filter out bad examples:

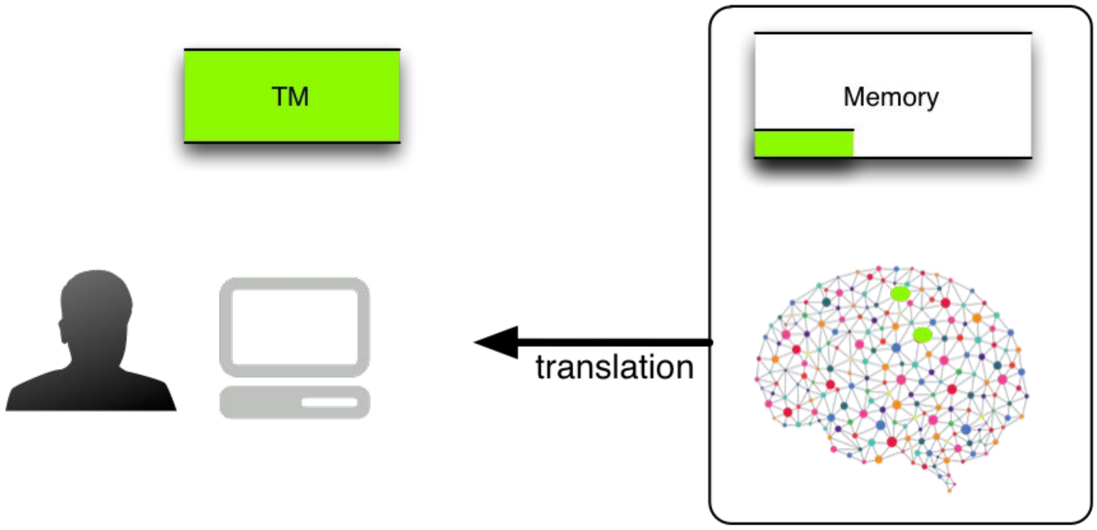
- Apply Fast-Align in two directions
- Compute Model 1 scores in two directions
- Combine and normalize scores
- Filter out on the distribution of scores

MODERN MT   Simple. Adaptive. Neural.

More Recent Adventures

Incremental Learning

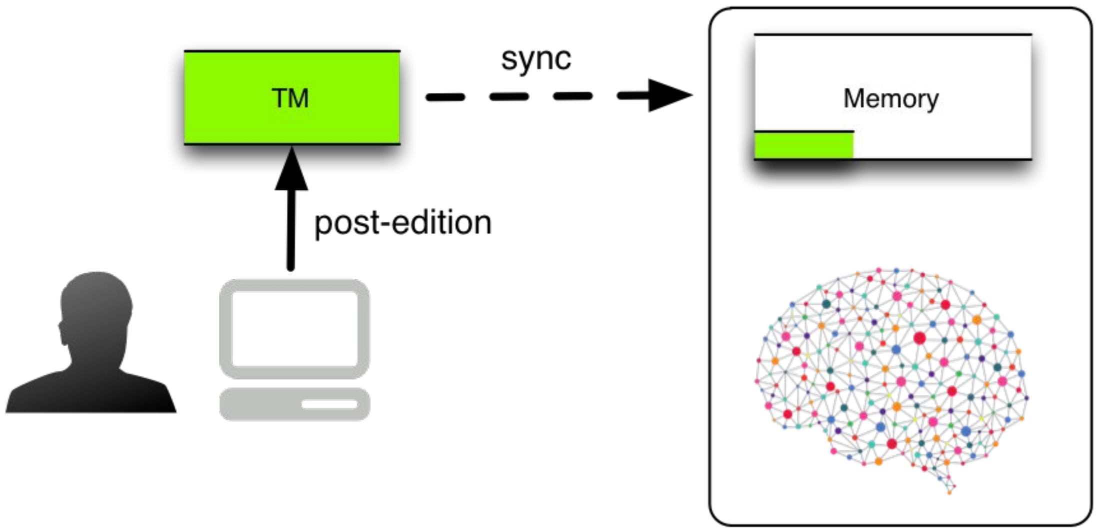
Incremental Learning



MODERN MT

FONDAZIONE BRUNO KESSLER

Incremental Learning



MODERN MT

FONDAZIONE BRUNO KESSLER

What happens when a new TM is uploaded?

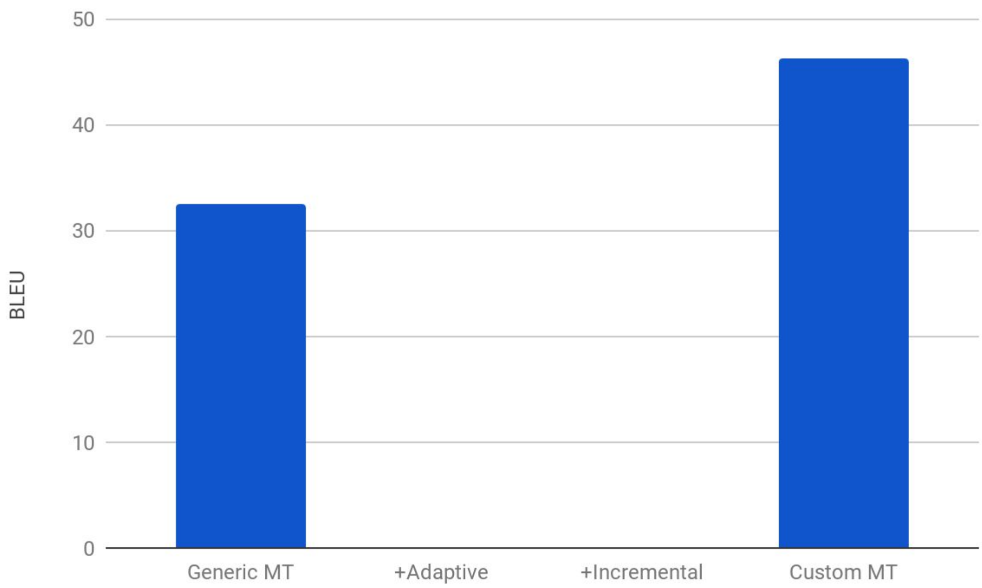
We compare:

- Generic MT: production engine [En-It]
- Custom MT: Generic MT tuned on TM [takes hours]
- +Adaptive MT: Generic MT adapted on TM [real-time]
- +Incremental MT: TM updated with simulated PE

MODERN MT

FONDAZIONE BRUNO KESSLER

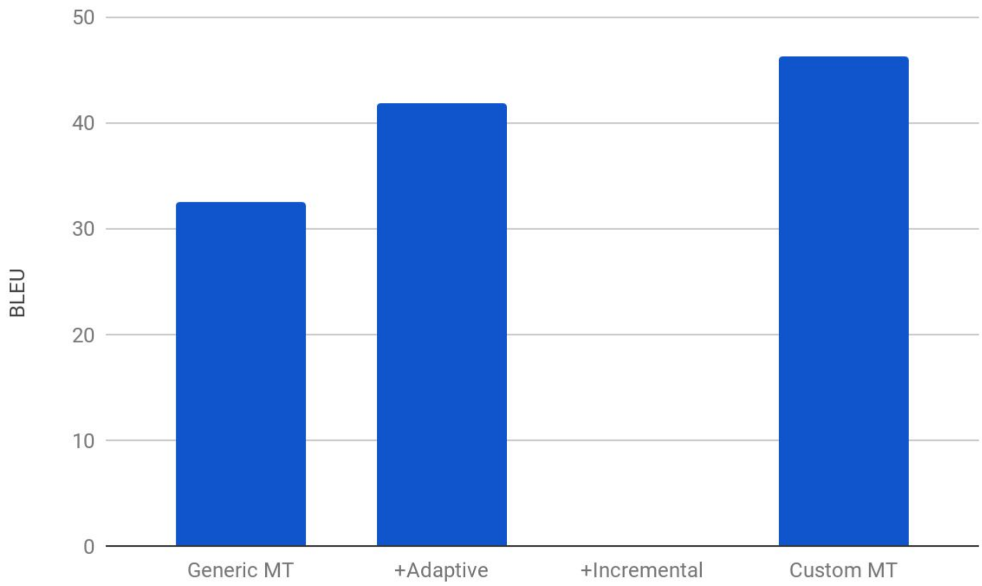
Incremental Learning



MODERN MT

FONDAZIONE BRUNO KESSLER

Incremental Learning

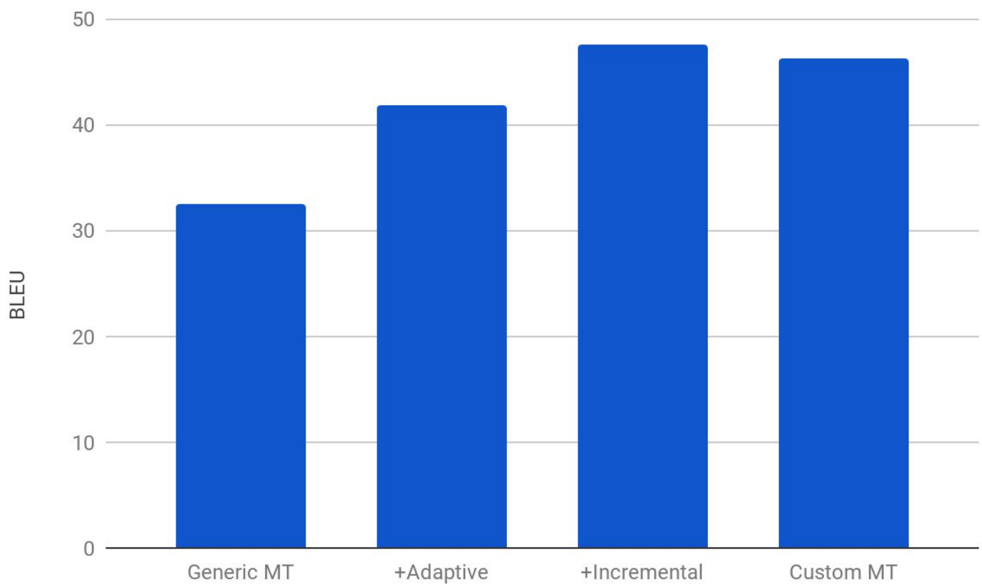


MODERN MT

FONDAZIONE BRUNO KESSLER

47

Incremental Learning

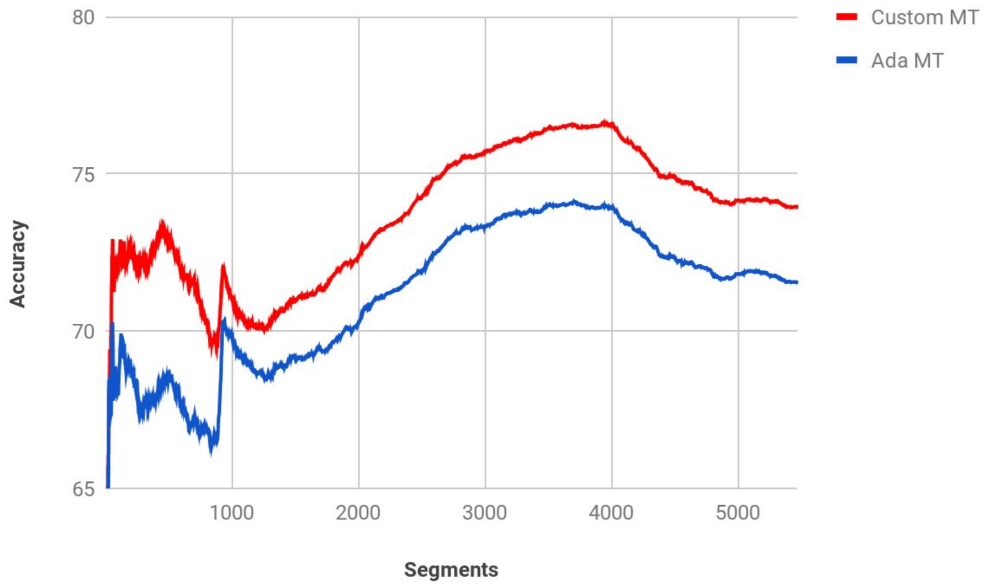


MODERN MT

FONDAZIONE BRUNO KESSLER

48

Incremental Learning

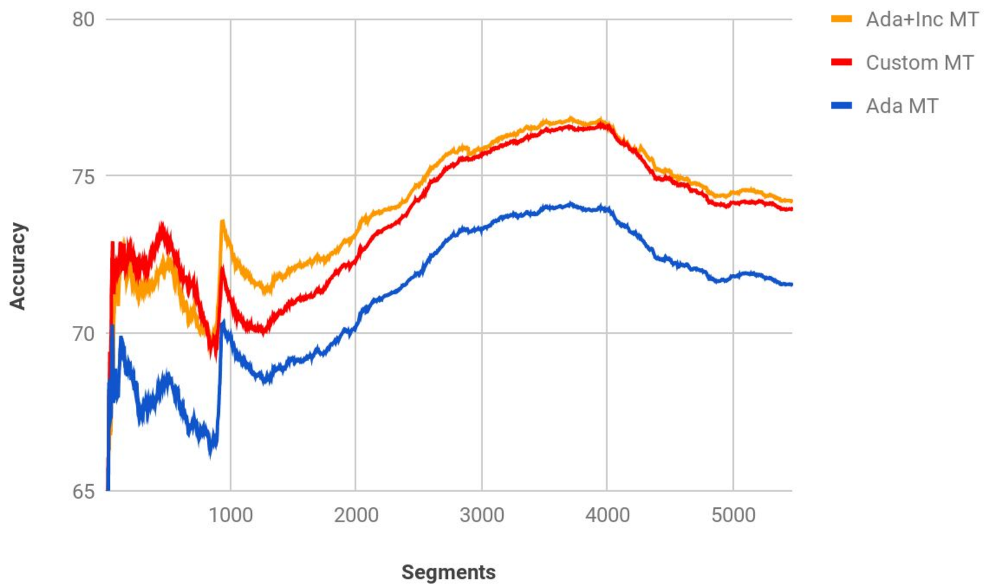


MODERN MT

FONDAZIONE BRUNO KESSLER

49

Incremental Learning



MODERN MT

FONDAZIONE BRUNO KESSLER

50

Online learning
(Adaptation *a posteriori*)

MODERN MT

FONDAZIONE BRUNO KESSLER

Online Learning

Use post-editing as new training instances

Perform one/more iterations

Can be combined with *a priori* adaptation

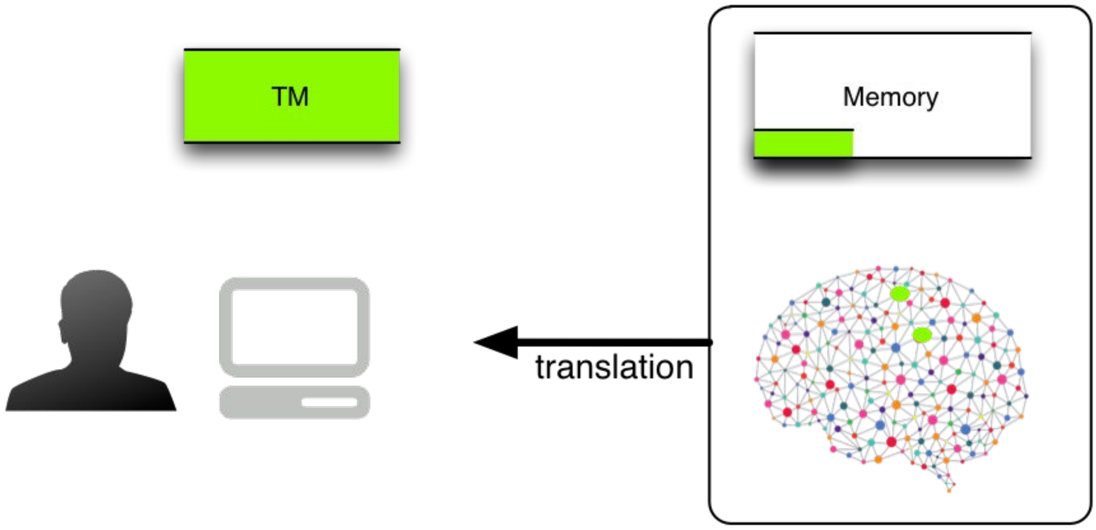
Updates generic or adapted model

Turchi et al. (2017), *Continuous learning from human post-edits for NMT, EAMT.*

MODERN MT

FONDAZIONE BRUNO KESSLER

Online Learning

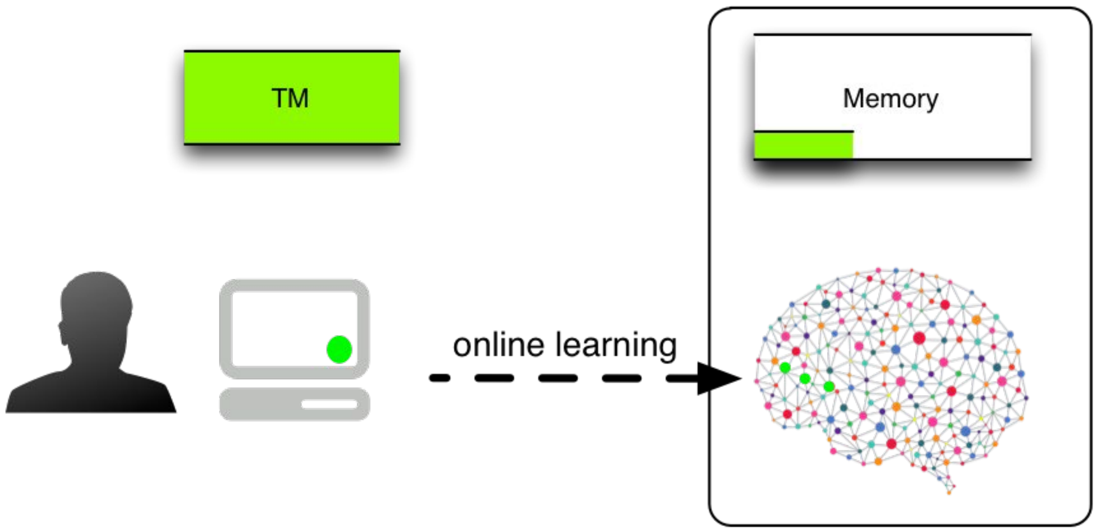


MODERN MT

FONDAZIONE BRUNO KESSLER

53

Online Learning

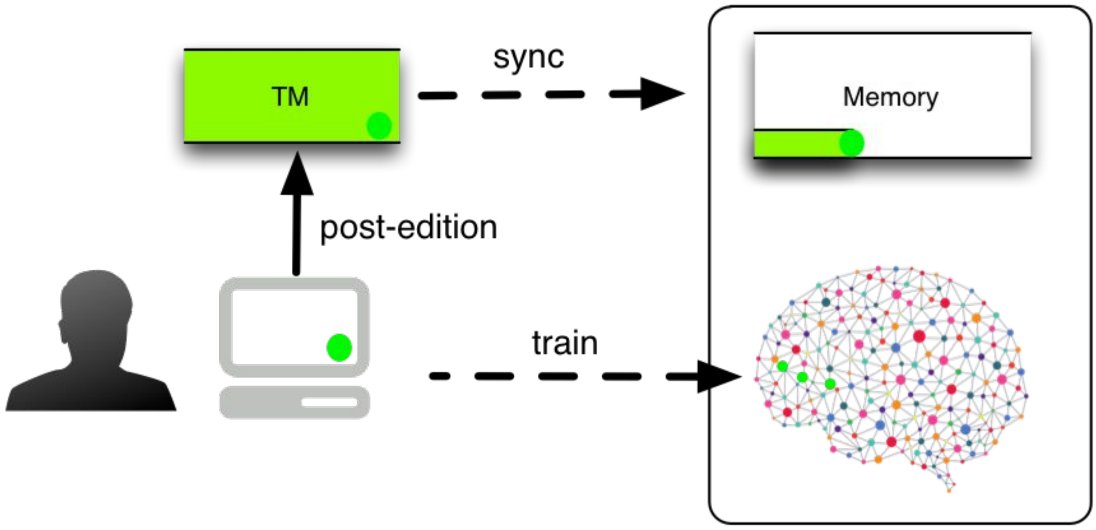


MODERN MT

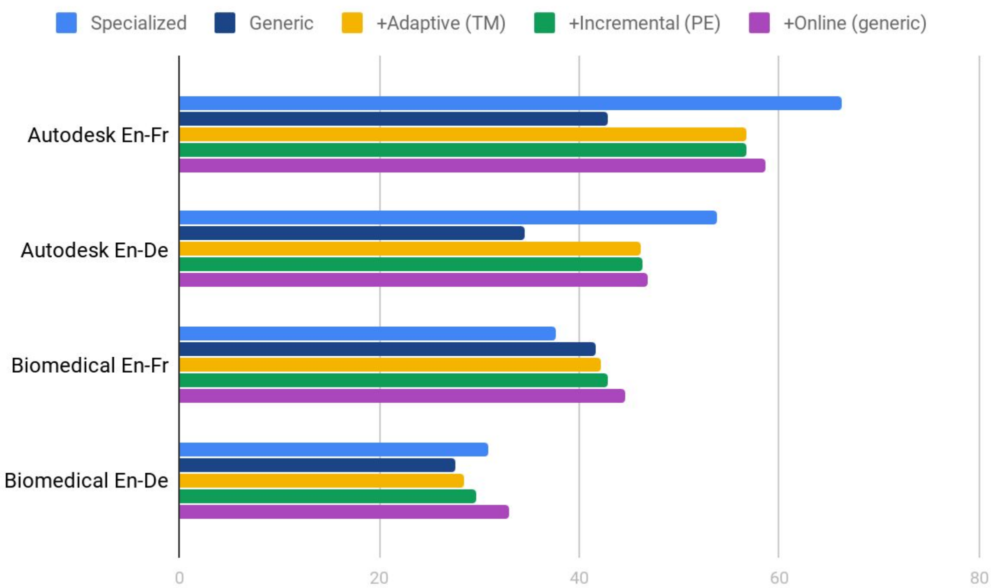
FONDAZIONE BRUNO KESSLER

54

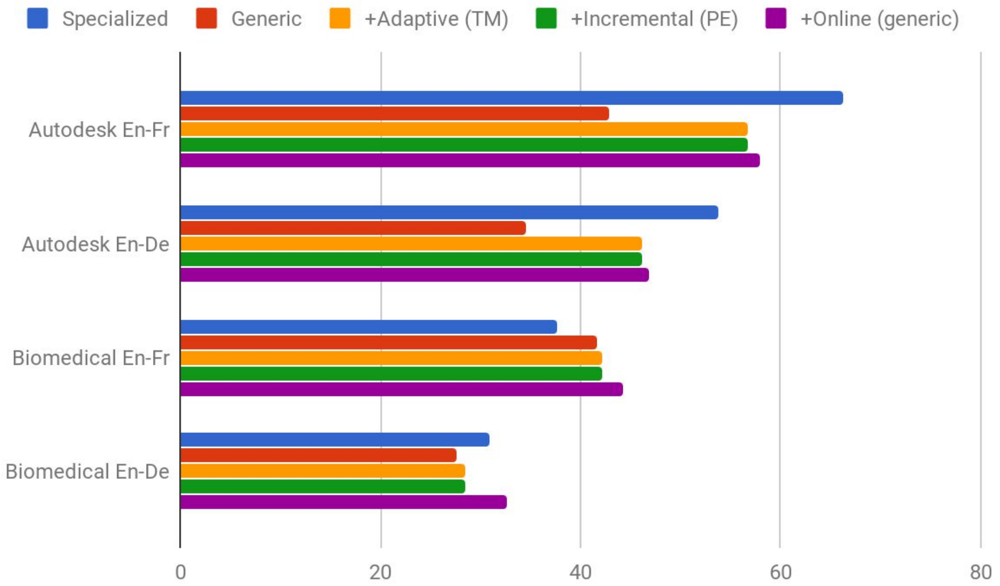
Incremental+Online Learning



Incremental+Online Learning (single domains)



Incremental+Online Learning (two domains)



MODERN MT

FONDAZIONE BRUNO KESSLER

57

Challenges

Online-learning contribution is consistent

Does it scale with number of domains?

Incremental learning contributes marginally

Probably depends on test set size

We are not always able to beat specialized models

How to improve further adaptation ?

MODERN MT

FONDAZIONE BRUNO KESSLER

58

Automatic Post-Editing

MODERN MT

FONDAZIONE BRUNO KESSLER

Automatic Post-Editing

Can improve MT without touching it inside

We can adapt an "external" MT service!

Similar to NMT: two inputs (*src*,*mt*), one output (*ape*)

Can be trained with less data than NMT

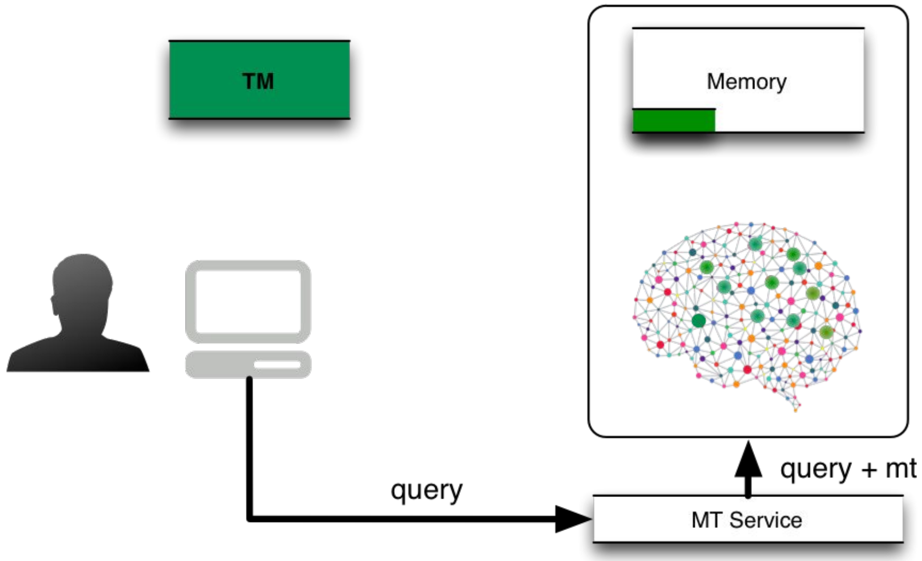
We can deploy instance based adaptation

Chatterjee et al. (2017), *Multi-source Neural APE: FBK's participation …. , WMT*.

MODERN MT

FONDAZIONE BRUNO KESSLER

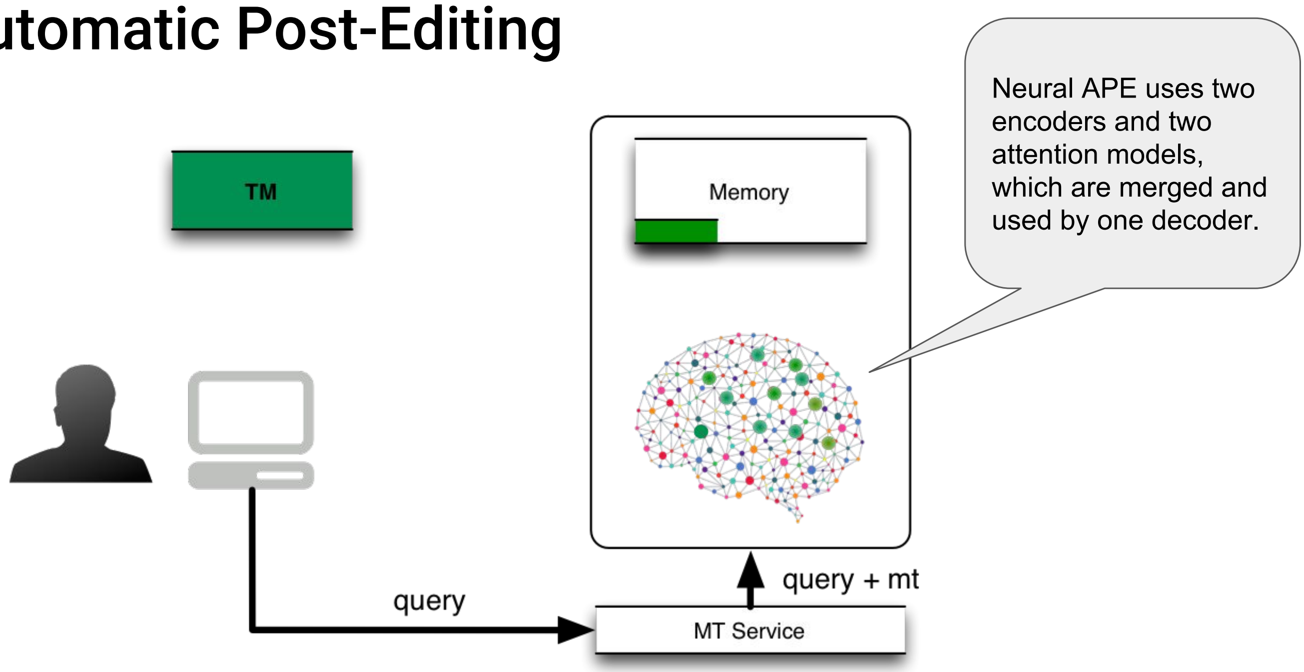
Automatic Post-Editing



MODERN MT

FONDAZIONE BRUNO KESSLER

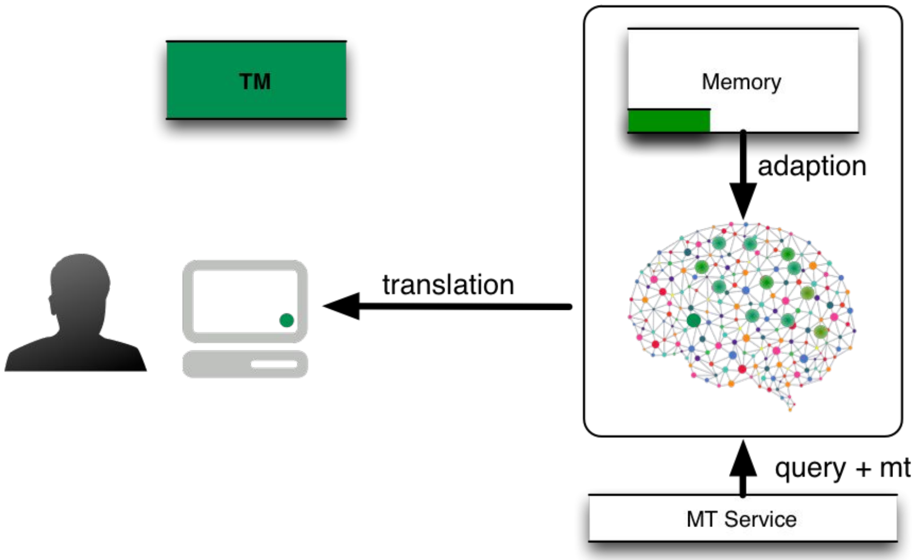
Automatic Post-Editing



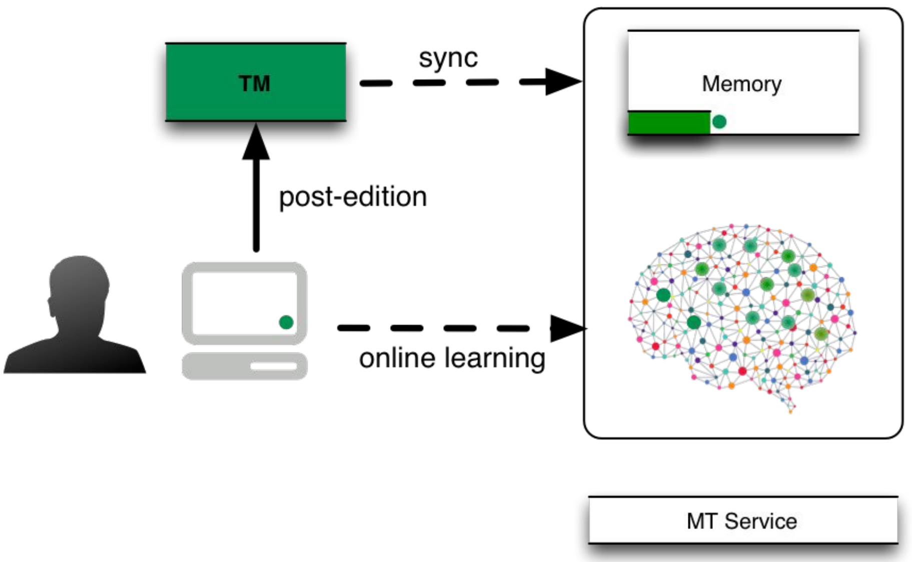
MODERN MT

FONDAZIONE BRUNO KESSLER

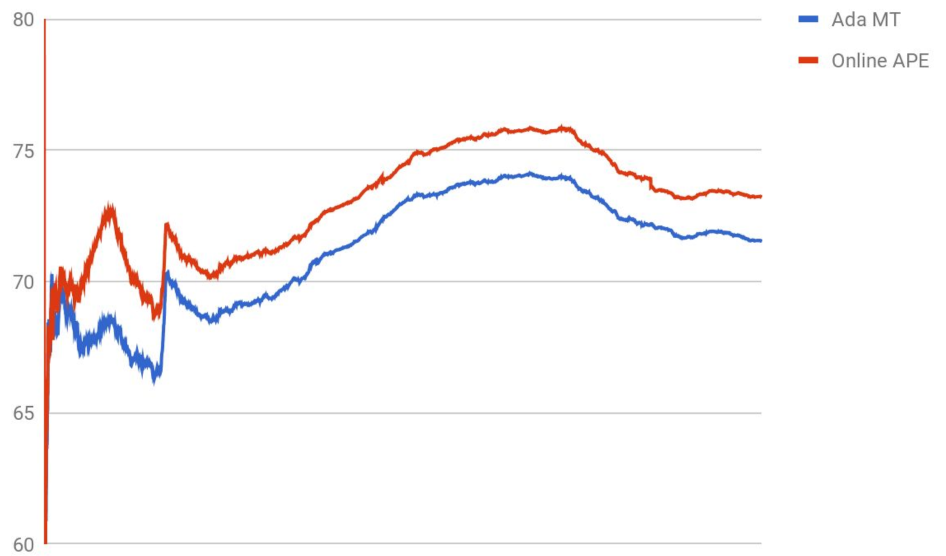
Automatic Post-Editing



Automatic Post-Editing



Automatic Post-Editing

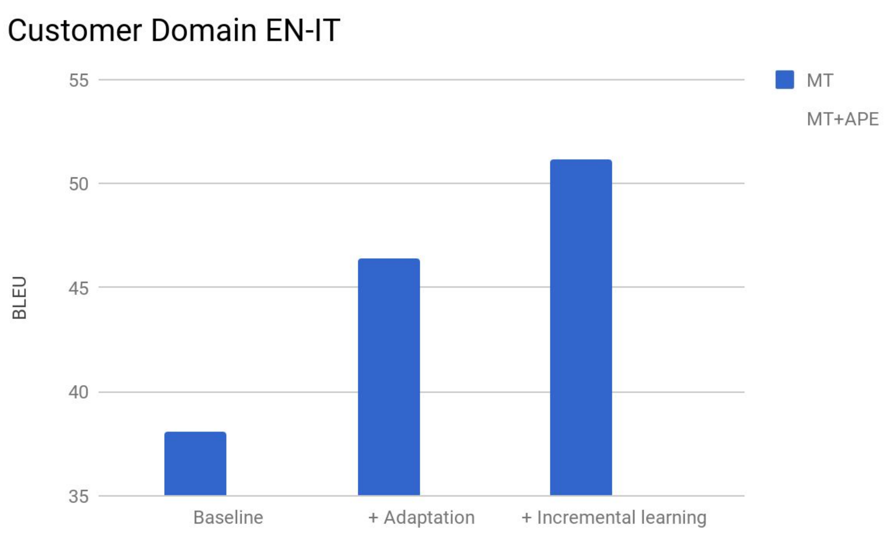


MODERN MT

FONDAZIONE BRUNO KESSLER

65

Automatic Post-Editing

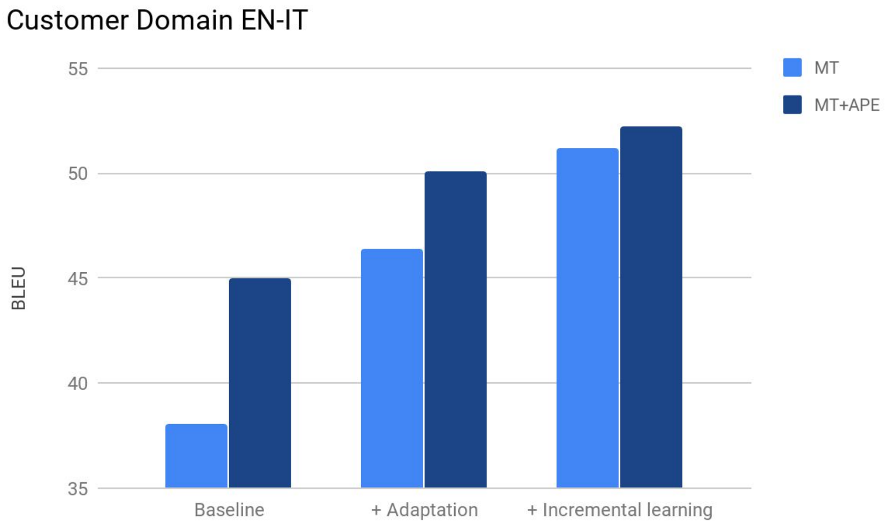


MODERN MT

FONDAZIONE BRUNO KESSLER

66

Automatic Post-Editing



MODERN MT

FONDAZIONE BRUNO KESSLER

Automatic Post-Editing

Can improve on top of static and adaptive engine!

Uses incremental learning, adaptation and online learning

Portable (in principle) on the multi-domain setting

Limited gain on top of full-fledged adaptive NMT

Can be an extra component to manage

MODERN MT

FONDAZIONE BRUNO KESSLER

Conclusions

MODERN MT

FONDAZIONE BRUNO KESSLER

Conclusions

Multi-user scenario goes beyond simple domain adaptation

We need to handle multiple evolving *domains*

Domain customization is not an option

Real-time adaptation/learning works!

But, there is still room for improvement!

MODERN MT

FONDAZIONE BRUNO KESSLER

Thank You

Website
www.ModernMT.eu

Github
github.com/ModernMT/MMT