UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

ebay

Can Neural Machine Translation be Improved with User Feedback?

Julia Kreutzer[1], Shahram Khadivi[3], Evgeny Matusov[3,4], Stefan Riezler[1,2]

[1]Computational Linguistics & [2]IWR, Heidelberg University, Germany
[3]eBay Inc., Aachen, Germany, [4]now AppTek

How to learn from weak feedback?

Improving MT with weak feedback

Learning from ...

- Online Feedback:
  - REINFORCE for SMT & NMT (Sokolov et al., 2016b; Kreutzer et al., 2017)
  - Advantage Actor Critic (Nguyen et al., 2017)
  - WMT Shared Task: Product titles (Sokolov et al., 2017)
- Offline Feedback:
  - Counterfactual Learning for SMT (Lawrence et al., 2017b,a)

Improving MT with weak feedback

Learning from simulated...

- Online Feedback:
  - REINFORCE for SMT & NMT (Sokolov et al., 2016b; Kreutzer et al., 2017)
  - Advantage Actor Critic (Nguyen et al., 2017)
  - WMT Shared Task: Product titles (Sokolov et al., 2017)
- Offline Feedback:
  - Counterfactual Learning for SMT (Lawrence et al., 2017b,a)

Goal: Improve NMT with offline bandit feedback from real users

Why? cheap and abundant source for MT adaptation

Explicit feedback: from stars to BLEU?

Collecting Explicit Feedback

⇒ 69,412 translated (en-es) item titles with 148k individual ratings

Bandit-to-Supervised Conversion

Bandit-to-Supervised Conversion: treats logged translations as references and ignores (or filters by) the feedback.

Pros:

- fine-grained feedback might be too noisy
- use standard supervised learning objectives (e.g. MLE or MRT)

Cons:

- discarding potentially useful information
- overconfidence in logged translations

Bandit-to-Supervised Conversion

Bandit-to-Supervised Conversion: treats logged translations as references and ignores (or filters by) the feedback.

Pros:

- fine-grained feedback might be too noisy
- use standard supervised learning objectives (e.g. MLE or MRT)

Cons:

- discarding potentially useful information
- overconfidence in logged translations

⇒ Simple, but effective with large data and domain gap.

Counterfactual Learning with Deterministic Logs

Learn from a log of user interactions with deterministic outputs of a historic MT system: $L = \{(\mathbf{x}^{(h)}, \mathbf{y}^{(h)}, \Delta(\mathbf{y}^{(h)}))\}_{h=1}^{H}$.

1. Self-normalized Deterministic Propensity Matching: (Lawrence et al., 2017b)

$$R^{\mathsf{DPM}}(\theta) = \frac{1}{H} \sum_{h=1}^{H} \Delta(\mathbf{y}^{(h)})\, \bar{p}_\theta(\mathbf{y}^{(h)}|\mathbf{x}^{(h)})$$

user feedback: average user star ratings scaled to $[0, 1]$
model probability: renormalized over current mini-batch

$\Rightarrow$ multiplicative control variate (Swaminathan and Joachims, 2015)

Counterfactual Learning with Reward Estimator

2. Doubly Controlled Estimation: (Lawrence et al., 2017b)

$$R^{\mathrm{DC}}(\theta) = \frac{1}{H}\sum_{h=1}^{H}\Bigg[\Big(\Delta(\mathbf{y}^{(h)}) - \hat{\Delta}_{\phi}(\mathbf{y}^{(h)})\Big)\,\bar{p}_{\theta}(\mathbf{y}^{(h)}|\mathbf{x}^{(h)}) + \sum_{\mathbf{y}\in\mathcal{S}(\mathbf{x}^{(h)})}\hat{\Delta}_{\phi}(\mathbf{y})\,p_{\theta}(\mathbf{y}|\mathbf{x}^{(h)})\Bigg]$$

estimated reward: regression model trained on logged rewards
sampled outputs: expected estimated reward

⇒ multiplicative and additive control variate (Dudík et al., 2011; Jiang and Li, 2016)

Experiments with Explicit Feedback

Model		Test BLEU	Test TER
Pre-trained BL		28.38	57.58
Counterfactual	DPM	28.19	57.80
	DPM-random	28.19	57.64
	DC	28.41	64.25
Bandit-to-Supervised		34.47	47.97

$\Rightarrow$ Bandit-to-supervised $\gg$ counterfactual learning

Analysis: why does explicit feedback fail?

Quality issues: noisy or even adversarial feedback, no normalization or user-specific handling of ratings.

Validation by experts:

1. star ratings
   - low inter-annotator agreement (Fleiss' $\kappa = 0.12$) for experts
   - no correlation of expert and user ratings (Spearman's $\rho = -0.05$)
2. agree/disagree with user ratings
   - moderate inter-annotator agreement ($\kappa = 0.45$)
   - majority votes agree with 42.3% of the ratings
   - mostly agree with high user ratings, disagree with low user ratings

Implicit Feedback: from user queries to BLEU!

Collecting Implicit Feedback

Embed the feedback collection into a "back-translation" CLIR task:

$$\text{query (es)} \xrightarrow[\text{translation}]{\text{query}} \text{query (en)} \xrightarrow{\text{search}} \text{title (en)} \xrightarrow[\text{translation}]{\text{item}} \text{title (es)}$$

- Store queries and titles when user clicks on translated title.
- Filter out tuples
  - where query translation fails: query (es) $==$ query (en)
  - where search fails: query (en) $\notin$ title (en).

$\Rightarrow$ **164,065** tuples of queries (es) and item titles (en+es)

Assumption: users are likely to be satisfied with item title translations that match their query.

Word-based matching for word in translation y_t and query $\mathbf{q}$:

$$\text{match}(y_t, \mathbf{q}) = \begin{cases} 1, & \text{if } y_t \in \mathbf{q} \\ 0, & \text{otherwise.} \end{cases}$$

Integrate into **minimum risk training**: (Shen et al., 2016)

$$R^{\text{W-MRT}}(\theta) = \sum_{s=1}^{S} \sum_{\tilde{\mathbf{y}} \in \mathcal{S}(\mathbf{x}^{(s)})} \prod_{t=1}^{T} \left[q_\theta^\alpha(\tilde{y}_t | \mathbf{x}^{(s)}, \tilde{\mathbf{y}}_{<t}) \, \text{match}(\tilde{y}_t, \mathbf{q}^{(s)}) \right].$$

Experiments with Implicit Feedback

Model	Test BLEU	Test TER	Test Query Recall
Pre-trained BL	28.38	57.58	45.96
Bandit-to-Supervised	34.39	47.94	63.21
Query Matching	34.52	46.91	68.12

$\Rightarrow$ Learning from query matching outperforms bandit-to-supervised

Summary

Learning from **explicit** feedback:

- counterfactual objectives failed (not in simulation!)
- bandit-to-supervised conversion is effective
- problematic noise level

Instead succeeded with **implicit** feedback:

- embed feedback collection in search task
- improvements in BLEU and task-specific metric

Questions?

How is reliability connected to learnability?

$\rightarrow$ ACL '18: Kreutzer et al. (2018)

How to learn from fine-grained feedback?

$\rightarrow$ ACL '18: Petrushkov et al. (2018); EAMT '18: Lam et al. (2018)

How to learn from human feedback for semantic parsing?

$\rightarrow$ ACL '18: Lawrence and Riezler (2018)

Thanks for your attention!

Additional Material

Distribution of User Ratings

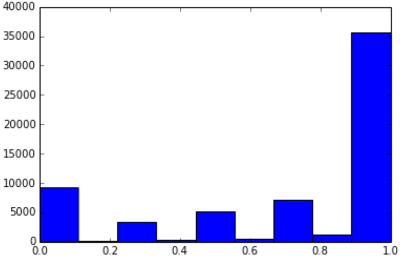


Figure 1: Average user ratings per item title (scaled).

Expert Validation Examples

Source:	Universal 4in1 Dual USB Car Charger Adapter Voltage DC 5V 3.1A Tester For iPhone
Translation:	Coche Cargador Adaptador De Voltaje Probador De Corriente Continua 5V 3.1A para iPhone
User Rating (avg):	4.5625
Expert Rating (avg):	4.33
User Judgment (majority):	Correct

Source:	BEAN BUSH THREE COLOURS: YELLOW BERGGOLD, PURPLE KING AND GREEN TOP CROP
Translation:	Bean Bush tres colores: Amarillo Berggold, púrpura y verde Top Crop King
User Rating (avg):	1.0
Expert Rating (avg):	4.66
User Judgment (majority):	Incorrect

Implicit Feedback Example

Query:	candado bicicleta
Translated Query:	bicycle lock
Title:	New Bicycle Vibration Code Moped Lock Bike Cycling Security Alarm Sound Lock
Translated Title:	Nuevo código de vibración Bicicleta Ciclomotor alarma de seguridad de bloqueo Bicicleta Ciclismo Cerradura De Sonido
Recall:	0.5

Objectives: MLE

Maximum Likelihood Estimation (MLE) on given parallel corpus of source and target sequences $D = \{(\mathbf{x}^{(s)}, \mathbf{y}^{(s)})\}_{s=1}^{S}$:

$$L^{MLE}(\theta) = \sum_{s=1}^{S} \log p_\theta(\mathbf{y}^{(s)}|\mathbf{x}^{(s)})$$

- requires references
- agnostic to rewards
- bandit-to-supervised conversion: use rewards to find translations to use as pseudo-references

Objectives: EL

Expected Loss (EL) maximizes the expectation of a reward over all source and target sequences: (Kreutzer et al., 2017; Sokolov et al., 2017)

$$R^{EL}(\theta) = \mathbb{E}_{p(\mathbf{x})p_\theta(\tilde{\mathbf{y}}|\mathbf{x})}\left[\Delta(\tilde{\mathbf{y}})\right]$$

- does not require references
- rewards are retrieved for sampled translations during learning
- in simulations: using smoothed sentence-level BLEU (sBLEU)

Objectives: MRT I

Minimum Risk Training (MRT) for NMT optimization from rewards if they can be obtained for several translations per input: (Shen et al., 2016)

$$R^{MRT}(\theta) = \sum_{s=1}^{S} \sum_{\tilde{\mathbf{y}} \in \mathcal{S}(\mathbf{x}^{(s)})} q_\theta^\alpha(\tilde{\mathbf{y}}|\mathbf{x}^{(s)})\, \Delta(\tilde{\mathbf{y}})$$

Sample probabilities are renormalized over a subset of translation samples $\mathcal{S}(\mathbf{x}) \subset \mathcal{Y}(\mathbf{x})$: $q_\theta^\alpha(\tilde{\mathbf{y}}|\mathbf{x}) = \frac{p_\theta(\tilde{\mathbf{y}}|\mathbf{x})^\alpha}{\sum_{\mathbf{y}' \in \mathcal{S}(\mathbf{x})} p_\theta(\mathbf{y}'|\mathbf{x})^\alpha}$.

Sequence-level rewards: all words of a translation are reinforced to the same extent and are treated as if they contributed equally to the translation quality.

Objectives: MRT II

Word-based rewards: allow the words to have individual weights

$$R^{W-MRT}(\theta) = \sum_{s=1}^{S} \sum_{\tilde{\mathbf{y}} \in \mathcal{S}(\mathbf{x}^{(s)})} \prod_{t=1}^{T} \left[q_\theta^\alpha(\tilde{y}_t | \mathbf{x}^{(s)}, \tilde{\mathbf{y}}_{<t}) \, \Delta(y_t) \right],$$

where $\Delta(y_t)$ is e.g. match with a given query.

Linear combination of MLE and (W)-MRT: (Wu et al., 2016)

$$R^{(W)-MIX}(\theta) = \lambda \cdot R^{MLE}(\theta) + R^{(W)-MRT}(\theta)$$

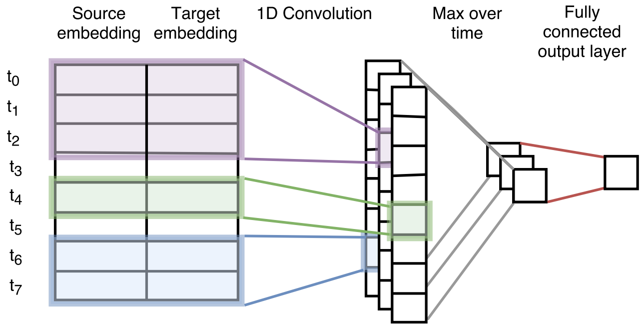
NMT Model

NMT baseline model:

- encoder-decoder architecture with attention (Bahdanau et al., 2015)
- sub-word vocabulary with byte-pair-encoding (Sennrich et al., 2016)
- eBay NMT: Python and TensorFlow
- trained with MLE on out-of-domain en-es parallel data
- beam of size 12 with length normalization (Wu et al., 2016)
- models with random sampling: avg & stdev for two runs

Learning from feedback starts from **pre-trained out-of-domain** baseline (=low-resource setting).

Reward Estimator



- embeddings are initialized with embeddings from BL NMT system and fine-tuned
- trained to minimize MSE on logged feedback

Reward Estimation Results

Data & Model	MSE	Macro-avg. Distance	Micro-Avg. Distance	Pearson's r	Spearman's ρ
Star ratings	0.1620	0.0065	0.3203	0.1240	0.1026
sBLEU	0.0096	0.0055	0.0710	0.8816	0.8675

Table 1: Results for the **reward estimators** trained and evaluated on human star ratings and simulated sBLEU.

Results on Public Data

Model	SMT	NMT (beam search)	NMT (greedy)
EP BL	25.27	27.55	26.32
NC BL	–	22.35	19.63
MLE	28.08	32.48	31.04
EL	–	28.02	27.93
DPM	26.24	27.54	26.36
DC	26.33	28.20	27.39

Table 2: BLEU results for simulation models evaluated on the News Commentary test set (`nc-test2007`) with beam search and greedy decoding. SMT results are from Lawrence et al. (2017b).

Simulation Results

Learning	Model	Test BLEU	Test TER
Pre-trained	BL	28.38	57.58
Fully Supervised	MLE	31.72	53.02
	MIX	34.79	48.56
Online Bandit	EL	31.78	51.11
Counterfactual	DPM	30.19	56.28
	DPM-random	28.20	57.89
	DC	31.11	55.05

Table 3: Results for **simulation experiments** evaluated on the product titles test set.

Translation Examples I

Title (en)	hall linvatec pro2070 powerpro ao drill synthes dhs & dcs attachment / warranty
Reference-0 (es)	hall linvatec pro2070 powerpro ao taladro synthes dhs & dcs accesorio / garantía
Reference-1 (es)	hall linvatec pro2070 powerpro synthes , perforación , accesorio de dhs y dcs , todo original , garantía
BL (es)	hall linvatec pro2070 powerpro ao perforadora synthes dhs & dcs adjuntos / garantía
MIX on star-rated titles (es)	hall linvatec pro2070 powerpro ao perforadora synthes dhs & dcs adjuntos / garantía
MIX on query-titles, small (es)	hall linvatec pro2070 powerpro ao perforadora synthes dhs & dcs adjuntos / garantía
MIX on query-titles, all (es)	hall linvatec pro2070 powerpro ao taladro synthes dhs & dcs adjuntos / garantía
W-MIX	hall linvatec pro2070 powerpro ao taladro synthes dhs & dcs accesorio / garantía

Translation Examples II

Title (en)	Unicorn Thread 12pcs Makeup Brushes Set Gorgeous Colorful Foundation Brush
Query (es) Query (en)	unicorn brushes // makeup brushes // brochas de unicornio // brochas unicornio unicorn brushes // makeup brushes
BL (es) Log (es)	galletas de maquillaje de 12pcs Unicorn Rosca 12 un. Conjunto de Pinceles para Maquillaje Hermosa Colorida Base Cepillo
W-MIX	unicornio rosca 12pcs brochas maquillaje conjunto precioso colorido fundación cepillo
Title (en)	12 × Men Women Plastic Shoe Boxes 33*20*12cm Storage Organisers Clear Large Boxes
Query (es) Query (en)	cajas plasticas para zapatos plastic shoe boxes
BL (es) Log (es)	12 × hombres mujeres zapatos de plástico cajas de almacenamiento 33*20*12cm organizadores de gran tamaño 12 × Zapato De Hombre Mujer De Plástico Cajas Organizadores de almacenamiento 33*20*12cm cajas Grande Claro
W-MIX	12 × para hombres zapatos de plástico cajas de plástico 33*20*12cm almacenamiento organizador transparente grandes cajas

References

References I

Bahdanau, D., Cho, K., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *Third International Conference on Learning Representations*, San Diego, California.

Dudík, M., Langford, J., and Li, L. (2011). Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on Machine Learning*, Bellevue, Washington.

Jiang, N. and Li, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, New York, NY.

References II

Kreutzer, J., Sokolov, A., and Riezler, S. (2017). Bandit structured prediction for neural sequence-to-sequence learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada.

Kreutzer, J., Uyheng, J., and Riezler, S. (2018). Reliability and learnability of human bandit feedback for sequence-to-sequence reinforcement learning. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, Melbourne, Australia.

Lam, T. K., Kreutzer, J., and Riezler, S. (2018). A reinforcement learning approach to interactive-predictive neural machine translation. In *Proceedings of the 21st Annual Conference of the European Association for Machine Translation (EAMT)*, Alicante, Spain.

References III

Lawrence, C., Gajane, P., and Riezler, S. (2017a). Counterfactual learning for machine translation: Degeneracies and solutions. In *Proceedings of the NIPS WhatIF Workshop*, Long Beach, CA.

Lawrence, C. and Riezler, S. (2018). Improving a neural semantic parser by counterfactual learning from human bandit feedback. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, Melbourne, Australia.

Lawrence, C., Sokolov, A., and Riezler, S. (2017b). Counterfactual learning from bandit feedback under deterministic logging : A case study in statistical machine translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark.

Nguyen, K., Daumé III, H., and Boyd-Graber, J. (2017). Reinforcement learning for bandit neural machine translation with simulated human feedback. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark.

Petrushkov, P., Khadivi, S., and Matusov, E. (2018). Learning from chunk-based feedback in neural machine translation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia.

Sennrich, R., Haddow, B., and Birch, A. (2016). Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, Germany.

References V

Shen, S., Cheng, Y., He, Z., He, W., Wu, H., Sun, M., and Liu, Y. (2016). Minimum risk training for neural machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, Germany.

Sokolov, A., Kreutzer, J., Lo, C., and Riezler, S. (2016a). Learning structured predictors from bandit feedback for interactive nlp. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, Germany.

Sokolov, A., Kreutzer, J., Riezler, S., and Lo, C. (2016b). Stochastic structured prediction under bandit feedback. In *Advances in Neural Information Processing Systems*, Barcelona, Spain.

References VI

Sokolov, A., Kreutzer, J., Sunderland, K., Danchenko, P., Szymaniak, W., Fürstenau, H., and Riezler, S. (2017). A shared task on bandit learning for machine translation. In *Proceedings of the Second Conference on Machine Translation*, Copenhagen, Denmark.

Swaminathan, A. and Joachims, T. (2015). The self-normalized estimator for counterfactual learning. In *Advances in Neural Information Processing Systems (NIPS)*, Montreal, Canada.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.