UNIVERSITAS
GADJAH MADA

Progress Report - Shared Task NLP

Deteksi Toxicity dalam Chat Game: Intent Detection and Slot Filling

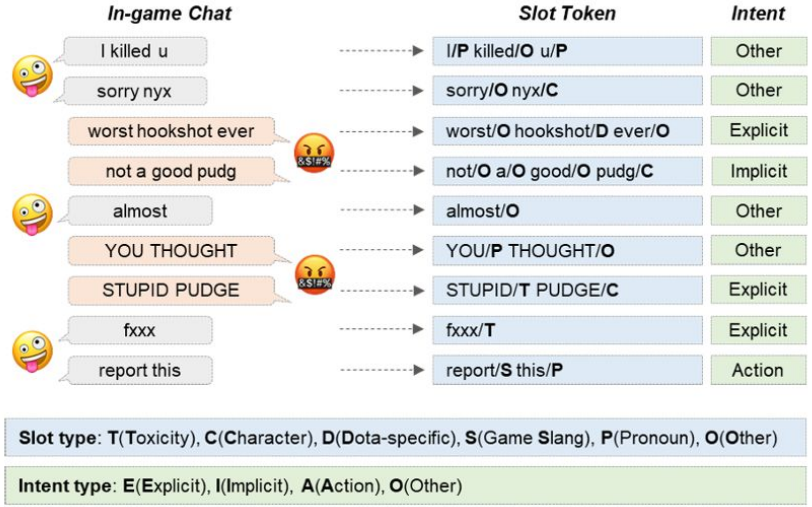
Bagus Cipta Pratama (23/516539/PA/22097)
Kosmas Rio Legowo (23/512012/PA/21863)
Muhammad Akmal Fauzan (23/519741/PA/22303)
Rafid Nur Huda (23/517734/PA/22205)

Penjelasan Task

Natural Language Understanding (NLU):

Shared task ini terdiri dari dua main task:

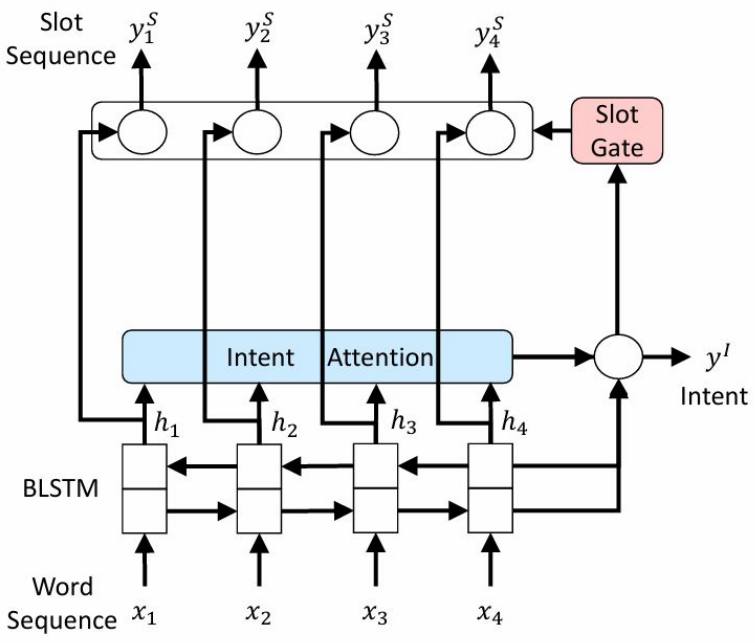
- Intent Classification → klasifikasi kalimat dari apa maksud (intent) chat pemain (**E**xplicit, **I**mplicit, **A**ction, **O**ther)
- Slot Filling → pelabelan urutan kata (level token) untuk mengambil informasi semantik yang ia bawa.
  (**T**oxic, **C**haracter, **D**ota-terms, game-**S**lang, **P**ronoun, **O**ther)



Arsitektur Model

- **Slot-Gated Joint NLU (Goo et al., 2018)**
- **Komponen Utama:**
  - **BiLSTM Encoder:** Memproses urutan kata dan menangkap dependensi kontekstual (Bidirectional LSTM).
  - **Intent Attention:** Menghasilkan intent context vector (c^I) sebagai representasi global kalimat.
  - **Slot Gate:** Menggabungkan informasi lokal (h_i) dan informasi intent (c^I) untuk mempengaruhi prediksi slot.
- **Joint Training:** Model dilatih bersamaan untuk prediksi *intent* dan *slot*.

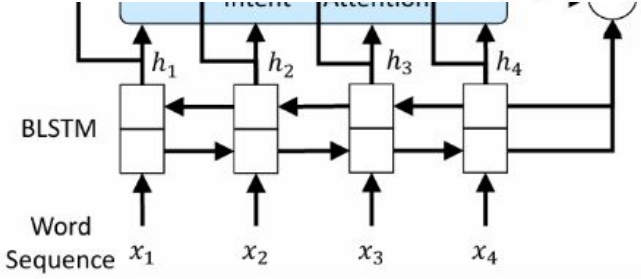
Arsitektur Slot-Gated Model (Intent Attention)



The workflow

- Input: Word Sequence, setiap kata akan diubah menjadi embedding vector .

$$(x_1, x_2, x_3, x_4)$$



- Encoder: BiLSTM .

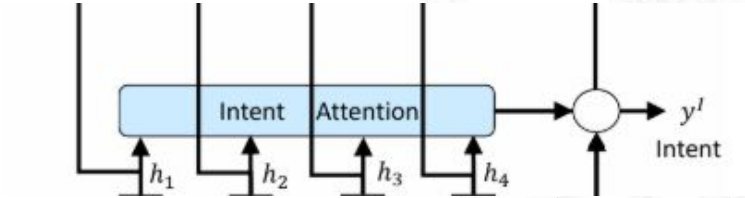
  Hidden state inilah representasi konteks kata ke-i.

$$h_i = [\overrightarrow{h_i}; \overleftarrow{h_i}]$$

The workflow

- Intent Attention
  Model menghitung context vector untuk intent dari seluruh hidden state

$$c^I = \sum_j \alpha_j h_j$$



  dimana a_j adalah attention weight terhadap kata j
- Setelah mendapat c^I, intent diprediksi

$$y^I = \text{softmax}(W(h_T + c^I))$$

Slot Gate

- menghubungkan pemahaman intent (global) dengan prediksi slot (per kata). gate dihitung sebagai

$$g_i = v^\top \tanh(h_i + W c^I)$$

Interpretasi:

- model membandingkan **representasi kata ke-i** dengan **informasi intent**
- gi sebesar apa menentukan berapa banyak informasi intent dipakai saat menebak slot kata itu

Jika gi besar: intent sangat relevan untuk slot token i

Jika gi kecil: slot lebih bergantung pada konteks kata itu sendiri

Slot Prediction

- Slot setiap kata diprediksi menggunakan

$$y_i^S = \text{softmax}(W(h_i + g_i \cdot h_i))$$

Arti dari (hi + gi · hi):

- Jika gi besar → signal intent “menguatkan” fitur slot
- Jika gi kecil → slot prediction lebih netral (hanya pakai hi

Sehingga model **secara adaptif** menggunakan intent jika memang relevan.

Joint Loss Function: Multi-Task Optimization Strategy

setelah mendapatkan keluaran berupa satu label per kalimat dan satu label per kata untuk token level , model dilatih dengan joint loss function

$$= \mathrm{CE}(y^I, \hat{y}^I) + \sum_{t=1}^{I} \mathrm{CE}(y_t^S, \hat{y}_t^S)$$

Cross-entropy di sini menjumlahkan hukuman untuk intent dan hukuman untuk setiap slot, sehingga kalau salah satu salah—bahkan satu token saja—loss langsung naik dan gradient memaksa semua parameter, termasuk gate, belajar dua tugas itu sekaligus.

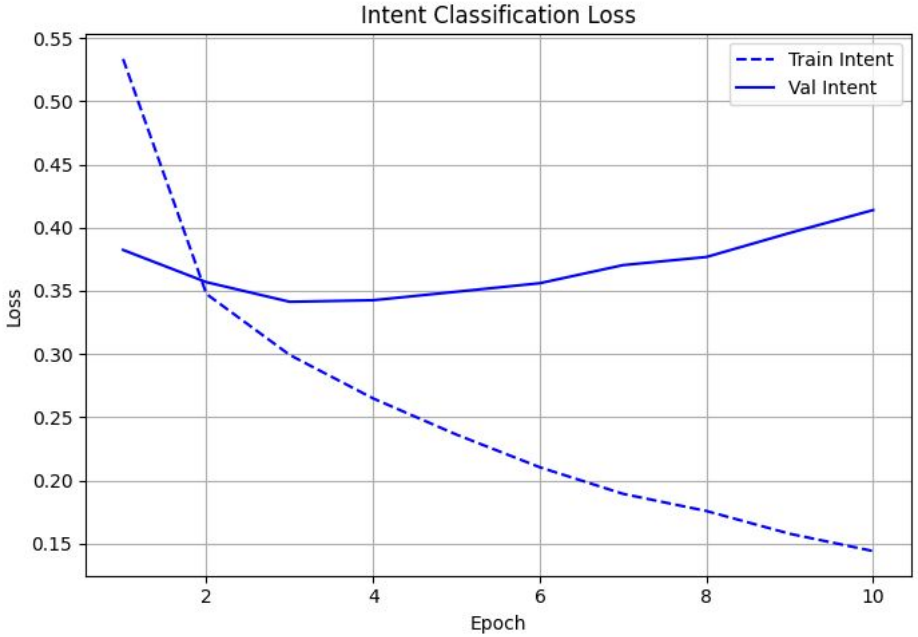
PARAMETER

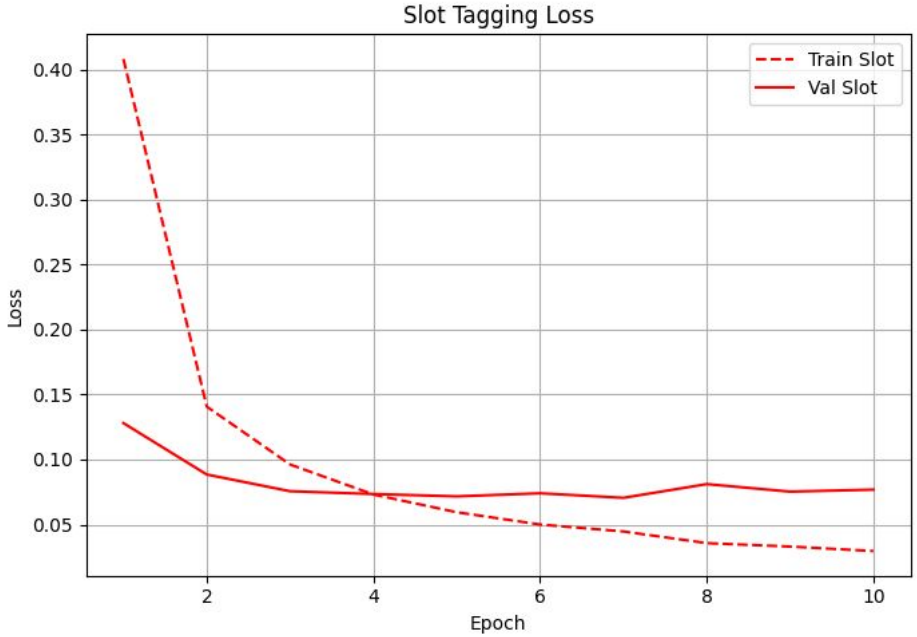
- **Epochs:** 10
- **Batch Size:** 32
- **Embedding dim:** 200
- **Learning rate:** 1e-3
- **Weight Decay:** 1e-4
- ***Hidden:*** *128*
- ***Dropout:*** *0.4*
- **Optimizer:** Adam *(learning rate = 1e-3)*

BASELINE VS OPTIMIZED

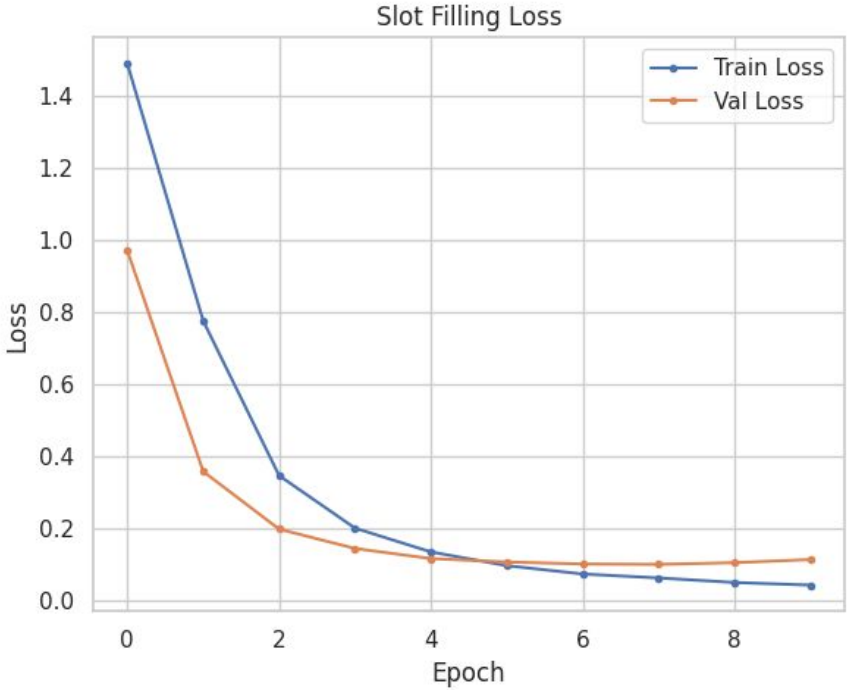
FITUR	BASELINE	OPTIMIZED
Input Embedding	Random Initialization	Word2Vec
Data handling	-	Data Augmentation (WordNet)
Loss Function	Standard Cross Entropy Loss	Weighted Cross Entropy Loss
Regularization	Dropout	Dropout dan Weight Decay

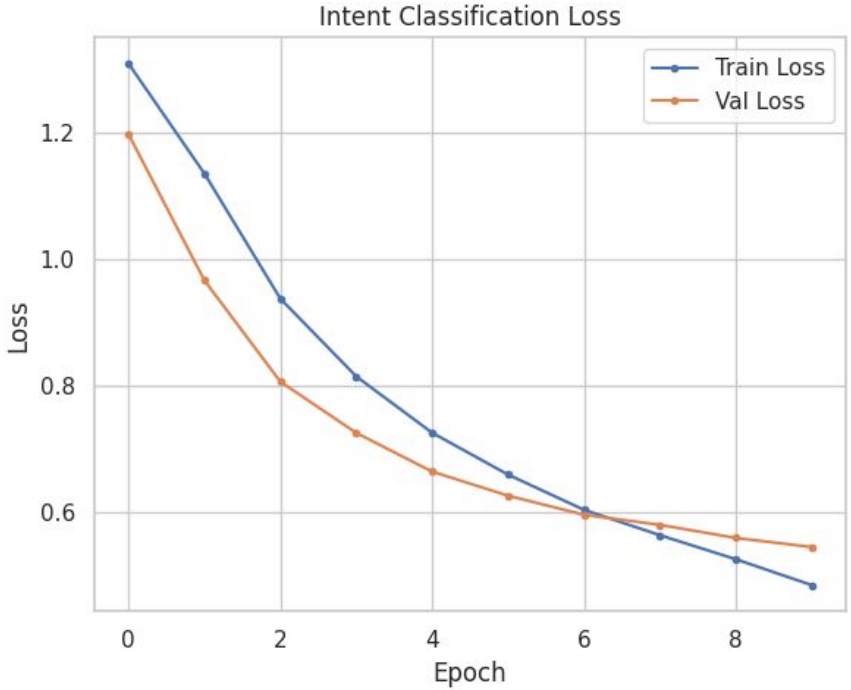
GRAFIK LOSS: BASELINE





GRAFIK LOSS: OPTIMIZED





Hasil Evaluasi pada Validation Set

- **Metrik yang Dilaporkan:**
  - **JSA (Joint Slot and Intent Accuracy):** Persentase kalimat dengan *intent* dan *slot* yang diprediksi benar.
  - **F1-score Intent:** 'E' (Explicit), 'I' (Implicit).
  - **F1-score Slot:** 'T' (Toxicity), 'D' (Dota-spesific), 'S' (game Slang).

Metric	Eksperimen 1	Eksperimen 2
JSA	0.8726	0.8523
F1(E)	0.83	0.83
F1(I)	0.72	0.65
F1(T)	0.96	0.96
F1(D)	0.93	0.94
F1(S)	0.99	0.99

References

Liu, B., & Lane, I. (2016). Attention-based recurrent neural network models for joint intent detection and slot filling. arXiv preprint arXiv:1609.01454.

Chih-Wen Goo, Guang Gao, Yun-Kai Hsu, Chih-Li Huo, Tsung-Chieh Chen, Keng-Wei Hsu, and Yun-Nung Chen. 2018. Slot-Gated Modeling for Joint Slot Filling and Intent Prediction. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), pages 753–757, New Orleans, Louisiana. Association for Computational Linguistics.

UNIVERSITAS
GADJAH MADA

Terima Kasih

matur suwun