



ArBanking77: Intent Detection Neural Model and a New Dataset in Modern and Dialectal Arabic

Mustafa Jarrar
Birzeit University, Palestine

Ahmet Birim
Sestek, Türkiye

Mohammed Khalilia
Birzeit University, Palestine

Mustafa Erden
Sestek, Türkiye

Sana Ghanem
Birzeit University, Palestine

Contributions

❖ ArBanking77 dataset

- ArBanking77 dataset consists of 31,404 queries.
- 2.4x larger than the Banking77 dataset.
- On average, there are 408 queries per intent
 - 202 MSA queries/intent
 - 206 Palestinian queries/intent.

❖ BERT Model

- F1-scores on MSA and PAL are 0.9209 and 0.8995, respectively

Used AraBBERT-V2 pre-trained model (Antoun et al., 2020)

ArBanking77 Dataset

The original Banking77 dataset

- 13,083 queries
- 77 classes (intents)
- Single domain, banking
- Open under the (CC-BY-4.0) license

The ArBanking77 corpus:

Each query in the original Banking77 has at least two corresponding queries in the ArBanking77

- At least one query written in MSA.
- At least one query written in Palestinian dialect.

Annotation Process

- 26 annotators (Well trained)
- Done using Google Sheets
- Over several months

Annotation Phases:

Phase 1: Arabization and Localization

Step1: The translation of the Banking77 from English into MSA.

- Done using Google Translate API.

Step2: The manual annotation .

The annotators performed four steps for each original English query:

- MSA_1 should be revised in case of incorrect translation.
- MSA_2 is optionally written by the annotator.
- PAL_1 is the formulation of the query in the Palestinian dialect.
- PAL_2 is optionally written by the annotator.

Each intent was divided among 2-5 annotators.

Phase 2: Review

Step1: Each annotator reviewed 3 belonged intents, to ensure that:

- The MSA and Palestinian queries should be acceptable, semantically correct and well-formulated.
- All queries in one intent belong to that intent, and not to other intents (labeling consistency).
- Spelling mistakes are ignored in order to simulate common errors and noise in real NLP systems, especially in live chat queries.

Step2: We revised duplicate queries by introducing additional variations to make them unique.

Lexical Relation between MSA and PAL

- Measured using the Jaccard Index for each parallel pair (MSA and PAL)

Results of Jaccard index:

- The mean is 0.16, the median 0.13, and the standard deviation 0.13.

Intent Detection Model

- BERT encoder is fine-tuned on Arabic intent detection task using the ArBanking77 dataset.

- A single linear layer was added on top of BERT transformer layers to perform the intent classification task.

Experiments and Results

• Zero-Shot Cross-Lingual Transfer Learning

- ❑ Used multi-lingual BERT (mBERT) (Devlinetal.,2018) and GigaBERT (Lanetal.,2020).

Result:

Pre-trained Model	Training Data	MSA F1	PAL F1
Multi-lingual BERT (uncased)	ArBanking77 (MSA)	-	0.5968
GigaBERT	Banking77 (English)	0.5047	0.3507
Multi-lingual BERT (uncased)	Banking77 (English)	0.1774	0.0903

• Pre-Trained Transformers Benchmark

- ❑ Evaluate various Arabic pre-trained transformer models, we benchmark against these models:

Results:

Pre-trained Model	MSA Test			PAL Test		
	Precision	Recall	F1	Precision	Recall	F1
AraBERTv2	0.9231	0.9212	0.9209	0.9004	0.9025	0.8995
MARBERTv2	0.9161	0.9142	0.9138	0.8983	0.8981	0.8962
ARBERT	0.9103	0.9121	0.9115	0.8810	0.8923	0.8899
QARIB	0.9147	0.9123	0.9121	0.8846	0.8864	0.8835
CAMEL-BERT-Mix	0.9149	0.9133	0.9128	0.8855	0.8854	0.8830
MARBERT	0.9106	0.9075	0.9070	0.8817	0.8817	0.8789
Multi-lingual BERT	0.8888	0.8872	0.8862	0.8598	0.8623	0.8578

• Low-Resource Simulation

- ❑ The size of samples one needs to achieve good and acceptable accuracy.

Results:

% of data	MSA Test			PAL Test		
	Precision	Recall	F1	Precision	Recall	F1
20%	0.8825	0.8755	0.8758	0.8441	0.8403	0.8363
50%	0.9117	0.9094	0.9088	0.8909	0.8903	0.8888
100%	0.9231	0.9212	0.9209	0.9004	0.9025	0.8995

• Noise and Error Simulation

- ❑ Experimented with three types of error and noise simulations:

- common spelling errors sim_c
- simulated errors sim_s
- keyboard-related errors sim_k

Results

Train Augmentation	Test Augmentation	MSA Test			PAL Test		
		20%	50%	100%	20%	50%	100%
None	None	0.8758	0.9088	0.9209	0.8363	0.8888	0.8995
	sim_c	0.8452	0.8795	0.8981	0.7933	0.8435	0.8637
	sim_s	0.8454	0.8813	0.8893	0.7585	0.8269	0.8463
	sim_k	0.8392	0.8648	0.8844	0.7942	0.8428	0.8634
sim_s/sim_k	None	0.8801	0.9126	0.9207	0.8421	0.8901	0.9018
	sim_c	0.8583	0.8922	0.9001	0.8065	0.8602	0.8711
	sim_s	0.8683	0.9017	0.9121	0.8055	0.8641	0.8857
	sim_k	0.8499	0.8833	0.8909	0.8086	0.8529	0.8749

Downloads and Demo

<https://sina.birzeit.edu/arbanning77/>

Public (data, code, demo)

