Distinguished organizing committee !

First of all, thanks to Chip2018 and Biendata competitions, it has been a pleasure for us to work on it, we learnt a lot of things, thank you !

We also thanks our team for their patience while we were coding day and night.

The following is our method description, you’re be free to correct:

**1/Features**

We distinguish two kind of features : embedding features, classical text mining features.

**Embedding features:**

Word vectors based on pre-training, we extracted the following distance features:

cosine\_distance,cityblock\_distance,jaccard\_distance,canberra\_distance,euclidean\_distance,minkowski\_distance,braycurtis\_distance,skew\_q1vec,skew\_q2vec,kur\_q1vec

**Classical text mining features:**

**Base word:**

q1\_word\_count, q2\_word\_count, word\_count\_diff, word\_overlap, uni\_BLEU, bi\_BLEU,BLEU2, char\_unigram\_overlap, q1\_hash, q2\_hash, q1\_freq, q2\_freq, word\_match, tfidf\_word\_match, min\_freq,common\_neighbours,fuzzy\_qratio,fuzzy\_wratio,fuzzy\_partial\_ratio,fuzzy\_partial\_token\_set\_ratio,fuzzy\_partial\_token\_sort\_ratio,fuzzy\_token\_set\_ratio,fuzzy\_token\_sort\_ratio

**Base char:**

q1\_char\_count, q2\_char\_count, char\_count\_diff, char\_overlap, uni\_BLEU, bi\_BLEU,BLEU2, q1\_hash,q2\_hash,q1\_freq,q2\_freq,char\_match,tfidf\_char\_match,min\_freq,common\_neighbours,fuzzy\_qratio,fuzzy\_wratio,fuzzy\_partial\_ratio,fuzzy\_partial\_token\_set\_ratio,fuzzy\_partial\_token\_sort\_ratio,fuzzy\_token\_set\_ratio,fuzzy\_token\_sort\_ratio

**2/Model**

We worked on main architectures for our Nets : A Decomposable Attention Neural Networks.

Decomposable attention(<https://arxiv.org/abs/1606.01933>) with pretrained word embedding.

Neural networks trained on both text sequences and our text mining features / embedding features proved to be our best single models.

We set validation\_split=0.1, eventually, our val\_accuracy achieve 0.8620, and F1\_measures achieve 0.8546 on testset.