This is just my scratch notes: there may be some errors and wrong explanations. I will try to fix them later on.

Distributional semantics = context

Sparse vector = only 1 and 0s

Dense vector= the highest possibility with the highest probable one

Corpus of text

Word2Vec = (max)2 words | center word | (max)2 words

$P(w_{t+j} | w_t ; theta)$ = product of w(t+j) and w(t) and likelihood theta

If negative likelihood minimize (minimization of errors) = maximizing predictive accuracy

v(w) = center word

u(w) = context word

We have 2 parameters to be optimized (u and v are 2 lookup tables)

vector similarity= dot product of outside word and context word

A = e to power of (vector similarity) = positive value of the vector similarity (to get the range between 0 and 1)

Normalization =   A / summation of total values ???

V = the total number of words in a vocabulary

P(a and b) = p(a) * p(b)

P(wt-2, wt-1, wt+1, wt+2|wt) = P(wt-1|wt) *….*P(wt+2|wt)

Optimization = finding the nearest point with derivation : gradient

Log and exponent are reverse functions and so they cancel out each other

Observed vector – expectation of the center = u(o) – Summation(x belongs to v) probability of (x | c) . u(x)

Similarly goes to calculating v

Then we get updated two vectors

Make average = ½ * [u(find word) + v (find word)]

1. Skip-grams = context ko center py p predict (better)
2. CBOW (continuous bags of words)= center ko context py p predict

Embedding(queen) – Embedding(king) + e

Using softmax is cost-inefficient because we need to compute the summation of total words again and again.

Additional efficiency in training:

1. Negative sampling

How to choose a good window size?

  Hyper-parameters : Setting an arbitrary parameters

    Choose the hypermeter which gives the best accuracy in validation test.

    Online hyper-parameters tuning

Cross-lingual Embedding

According to the morphological problem, it is inappropriate to use word2vector. (Burmese)

  Broken Language, Spelling error

It is not okay to depend on the dictionary in Myanmar.

Word vector -> Sub-word vector level

BPE – Byte Pair Encoding