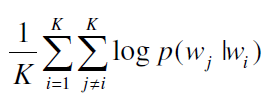
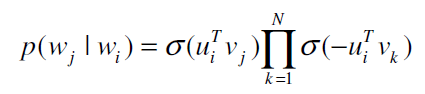
**Method**

Traditional Collaborative Filtering (CF) algorithms are item-based and they analyze item-item relations in order to produce item similarities. Skip-gram with Negative Sampling (SGNS), also known as Word2Vec, is a neural word embedding method introduced by Mikolov et. al [1]. It was shown to provide state-of-the-art results on various linguistics tasks. The method aims at finding words representation that captures the relation between a word to its surrounding words in a sentence. Inspired by several works in the field Natural Language Processing (NLP) that learned a latent representation of words using neural embedding algorithms, especially SGNS, a method named Item2Vec for item-based CF that produced embedding for items in a latent space was developed in 2016 [2]. The method can infer item-to-item relations even when user information is not available.

In the context of CF data, the items are given as user generated sets. First, the user's click list of the item is obtained, and the item set of the user u's evaluation is set as Set= {*i1*, *i2*, *i3*...}. Then, according to the skip-gram, the set of items is treated as a sentence (the spatial/time information is discarded). Each pair of items that share the same set is treated as a positive example. The window size is the determined from the set size.

The Skip-gram objective aims at maximizing the following term:

 (1)

where *K* is the length of the item list. Negative sampling comes to alleviate the computational problem by the replacement of the softmax function (2)

where  are latent vectors that correspond to the target and context representations for the item, respectively,, *N* is a parameter that determines the number of negative examples to be drawn per a positive example.

**Experimental Results**

Systems and parameters

We applied Item2Vec to both datasets. The optimization is done by stochastic gradient decent. We ran the algorithm for 60 epochs. We set *min\_count* to 1, so all the items can be keeped. The window size is 999999, much more than the list of the item lists. We use skip-gram so *sg* = 1, and we apply negative sampling so *hs* = 0. The dimension parameter *size* was set to 128. We set *seed* for random number generator to 42, and set *alpha* for learning rate to 0.025.【这个学习率是最影响结果的】

|  |  |  |
| --- | --- | --- |
| Parameters | Description | Value |
| iter | Number of iterations (epochs) over the corpus. | 60 |
| alpha | The initial learning rate. | 0.025 |
| seed | Seed for the random number generator | 42 |
| min\_count | Ignores all words with total frequency lower than this. | 1 |
| window | Maximum distance between the current and predicted word within a sentence. | 99999 |
| sg | Training algorithm: 1 for skip-gram; otherwise CBOW. | 1 |
| hs | If 1, hierarchical softmax will be used for model training. If 0, and negative is non-zero, negative sampling will be used. | 0 |
| **size** | Dimensionality of the word vectors. | 128 |

[1] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems 2013 (pp. 3111-3119).

[2] Barkan O, Koenigstein N. Item2Vec: Neural item embedding for collaborative filtering. In IEEE 26th International Workshop on Machine Learning for Signal Processing. 2016.