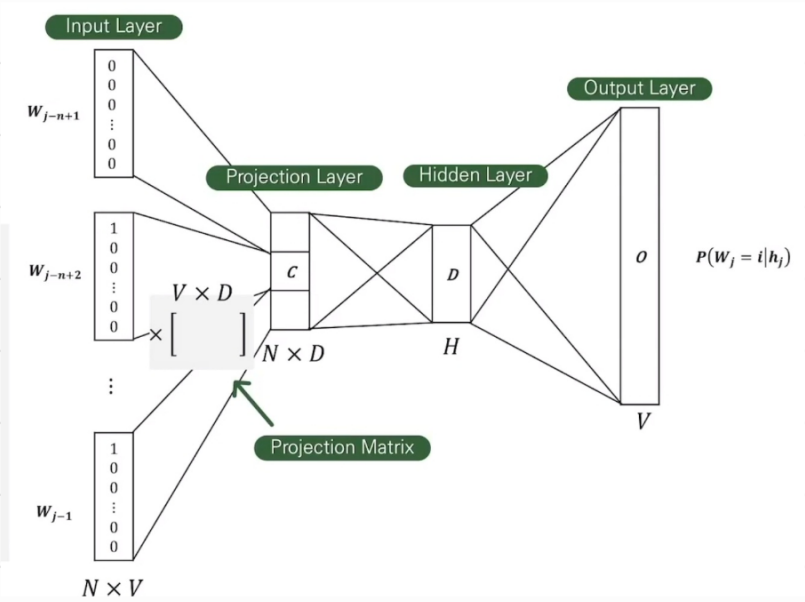
2. 기존에 문제를 풀었던 방법 ; NNLM



N : previous word의 개수

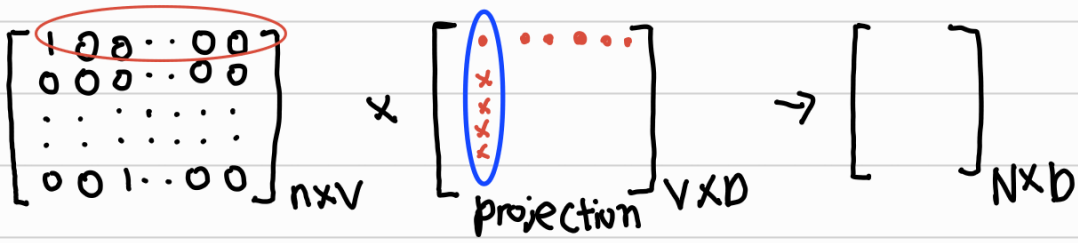
V : 전체 단어 수

N개의 단어들은 one-hot encoding으로 표현

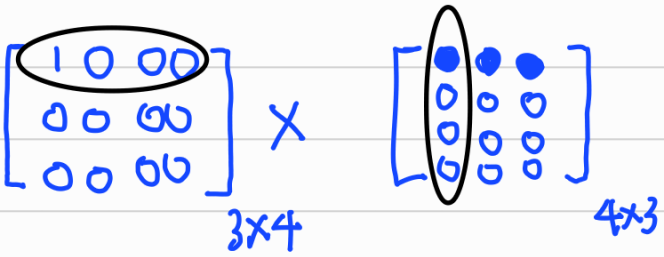
Q: projection layer 와 hidden layer 차이점?

activation function X activation function O

Input layer → projection layer



NxD 번만 계산하면 됨



→ 9번 계산

projection layer → hidden layer

PxH (about 2500 ~ 200000)

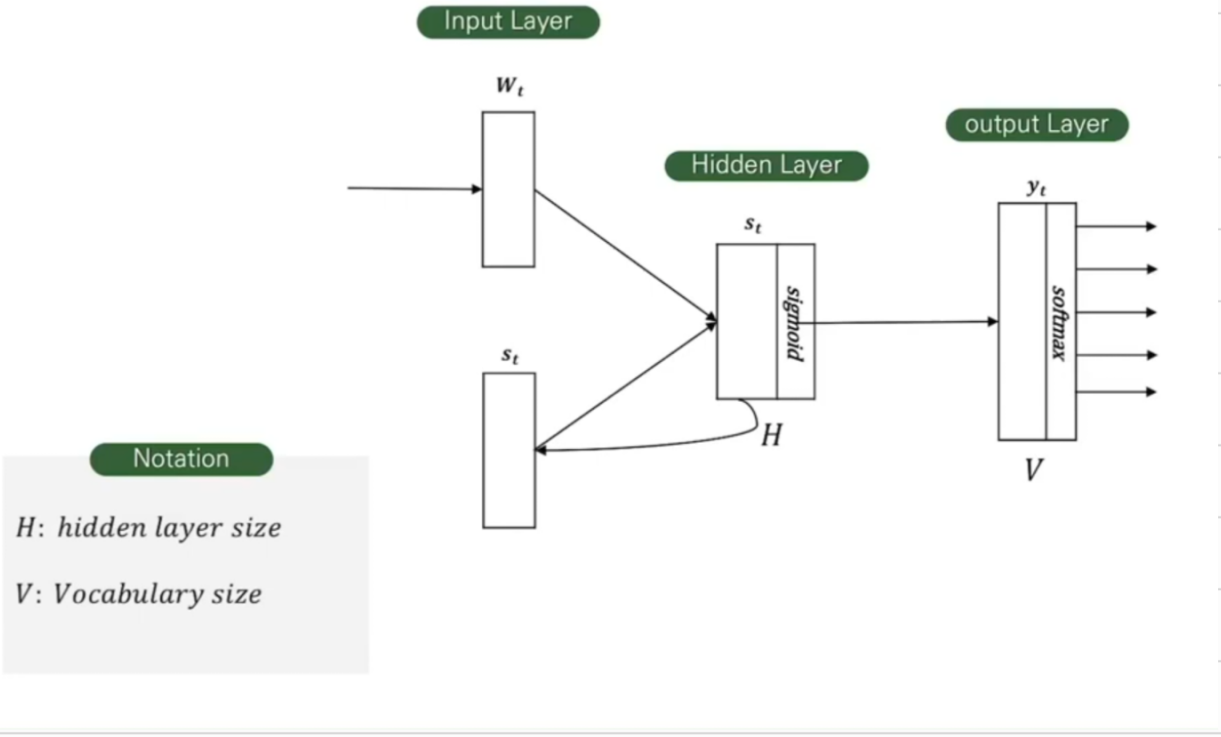
hidden layer → output layer

$H \times V \Rightarrow H \times \log_2 V$ (binary tree 이용)

최종 복잡도 : $N \times D + N \times D \times H + H \times \log_2 V$

2. 기존에 문제를 풀었던 방법 ; RNNLM

- Recurrent Neural Net Language Model (RNNLM)
  - ✓ Structure



NNLM의 한계: N의 크기를 정해줘야함

$h_t \leftarrow h_{t-1}, x_t$ // short-term 메모리의 역할

복잡도 : $H \times H + H \times V$

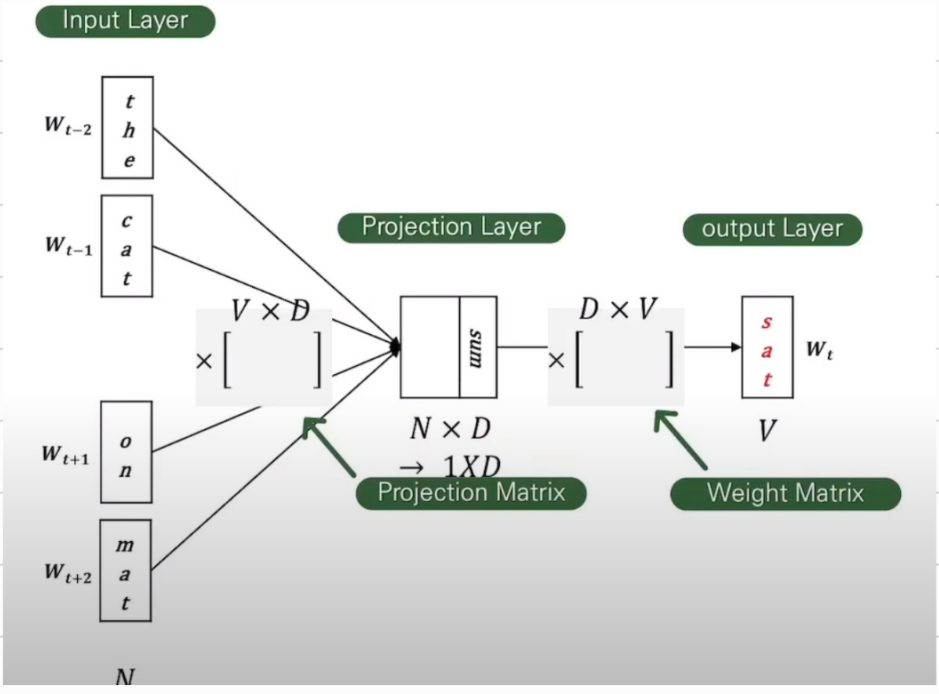
↳ $H \times H + H \times \log_2 V$

3. 논문에서 제시한 아이디어

NNLM의 경우: hidden layer에서 복잡도↑

⇒ hidden layers 삭제

4. 구현 : CBOW



Input layer → projection layer

NNLM 과 구조가 똑같음 ⇒ 복잡도 : N×D

projection layer → hidden layer

projection layers에서 모든 단어들의 벡터들이 합쳐짐. (averaged)

⇒ 단어들의 순서가 projection에 영향을 미치지 않음 (bag-of-words 모델)

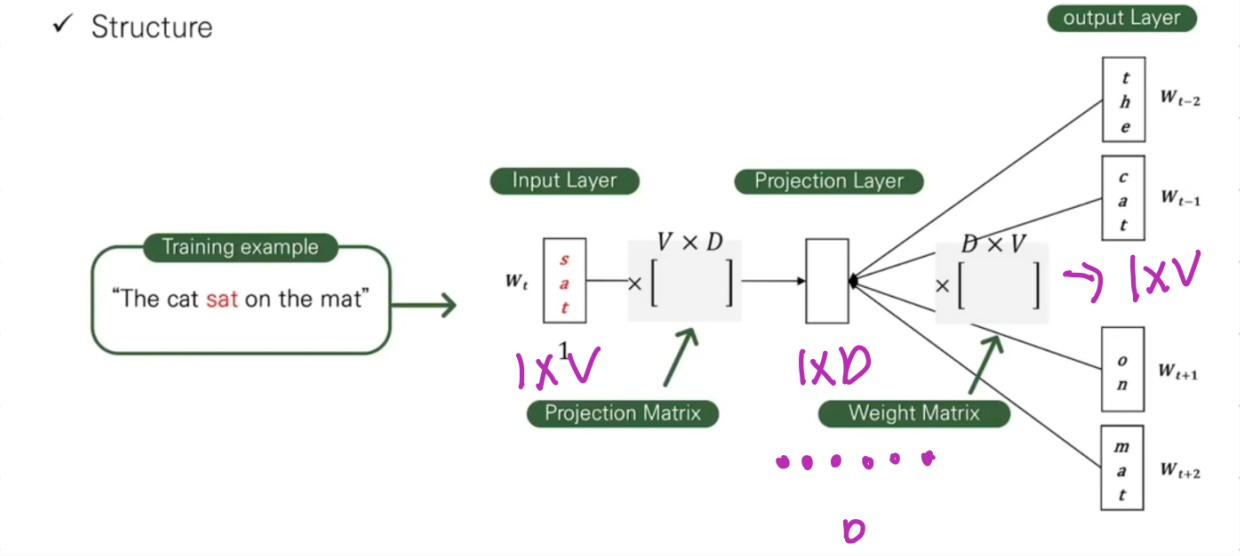
continuous distributed representation 를 써서 CBOW라 부름

복잡도 : D×V ⇒ D× $\log_2 V$

최종 복잡도 : N×D + D× $\log_2 V$

4. 구현 : Skip-gram

- Continuous Skip-gram Model
  - ✓ Structure



Input layer → projection layer

복잡도 : 1xD

projection layer → output layer

복잡도 : DxV ⇒ Dx $\log_2$V

이 짓을 C번 반복

C는 중심단어에서 최대 거리

최종 복잡도 : Cx(D+Dx $\log_2$V)

cf) training 할때 멀리 있는 단어는 less related

⇒ Sampling 조금만

how?

<1, C>에서 무작위 숫자 선택하고, 이를 R

Let C=5, R=3

training sentence : For 〈 writers, a random sent ence can help them get their creative 〉 juices flowing

5. 실험 및 결과

Syntatic : X = V(biggest) − V(big) + V(small)

Semantic : X = V(France) − V(Paris) + V(Berlin)

X를 잘 맞춤

Type of relationship	Word Pair 1		Word Pair 2	
Common capital city	Athens	Greece	Oslo	Norway
All capital cities	Astana	Kazakhstan	Harare	Zimbabwe
Currency	Angola	kwanza	Iran	rial
City-in-state	Chicago	Illinois	Stockton	California
Man-Woman	brother	sister	grandson	granddaughter
Adjective to adverb	apparent	apparently	rapid	rapidly
Opposite	possibly	impossibly	ethical	unethical
Comparative	great	greater	tough	tougher
Superlative	easy	easiest	lucky	luckiest
Present Participle	think	thinking	read	reading
Nationality adjective	Switzerland	Swiss	Cambodia	Cambodian
Past tense	walking	walked	swimming	swam
Plural nouns	mouse	mice	dollar	dollars
Plural verbs	work	works	speak	speaks

동의어 → 오답처리 100% 정확도 불가능

dimensionality와 Training words 둘 다↑ ⇒ accuracy↑

Table 2: *Accuracy on subset of the Semantic-Syntactic Word Relationship test set, using word vectors from the CBOW architecture with limited vocabulary. Only questions containing words from the most frequent 30k words are used.*

Dimensionality / Training words	24M	49M	98M	196M	391M	783M
50	13.4	15.7	18.6	19.1	22.5	23.2
100	19.4	23.1	27.8	28.7	33.4	32.2
300	23.2	29.2	35.3	38.6	43.7	45.9
600	24.0	30.1	36.5	40.8	46.6	50.4

CBOW, Skip-gram, NNLM, RNNLM 비교

Table 3: *Comparison of architectures using models trained on the same data, with 640-dimensional word vectors. The accuracies are reported on our Semantic-Syntactic Word Relationship test set, and on the syntactic relationship test set of [20]*

Model Architecture	Semantic-Syntactic Word Relationship test set		MSR Word Relatedness Test Set [20]
	Semantic Accuracy [%]	Syntactic Accuracy [%]	
RNNLM	9	36	35
NNLM	23	53	47
CBOW	24	64	61
Skip-gram	55	59	56

Table 5: *Comparison of models trained for three epochs on the same data and models trained for one epoch. Accuracy is reported on the full Semantic-Syntactic data set.*

Model	Vector Dimensionality	Training words	Accuracy [%]			Training time [days]
			Semantic	Syntactic	Total	
3 epoch CBOW	300	783M	15.5	53.1	36.1	1
3 epoch Skip-gram	300	783M	50.0	55.9	53.3	3
1 epoch CBOW	300	783M	13.8	49.9	33.6	0.3
1 epoch CBOW	300	1.6B	16.1	52.6	36.1	0.6
1 epoch CBOW	600	783M	15.4	53.3	36.2	0.7
1 epoch Skip-gram	300	783M	45.6	52.2	49.2	1
1 epoch Skip-gram	300	1.6B	52.2	55.1	53.8	2
1 epoch Skip-gram	600	783M	56.7	54.5	55.5	2.5

Table 7: *Comparison and combination of models on the Microsoft Sentence Completion Challenge.*

Architecture	Accuracy [%]
4-gram [32]	39
Average LSA similarity [32]	49
Log-bilinear model [24]	54.8
RNNLMs [19]	55.4
Skip-gram	48.0
Skip-gram + RNNLMs	**58.9**