Semantic Compositionality through Recursive Matrix-Vector Spaces

COURSE PROJECT OF CS365A

SONU AGARWAL

VIVEKA KULHARIA

Goal

- Classifying semantic relationships such as "cause-effect" or "component-whole" between nouns

  - Examples:

    - "The introduction in the book is a summary of what is in the text."
      - Component-Whole

    - "The radiation from the atomic bomb explosion is a typical acute radiation."
      - Cause-Effect

Parse Tree

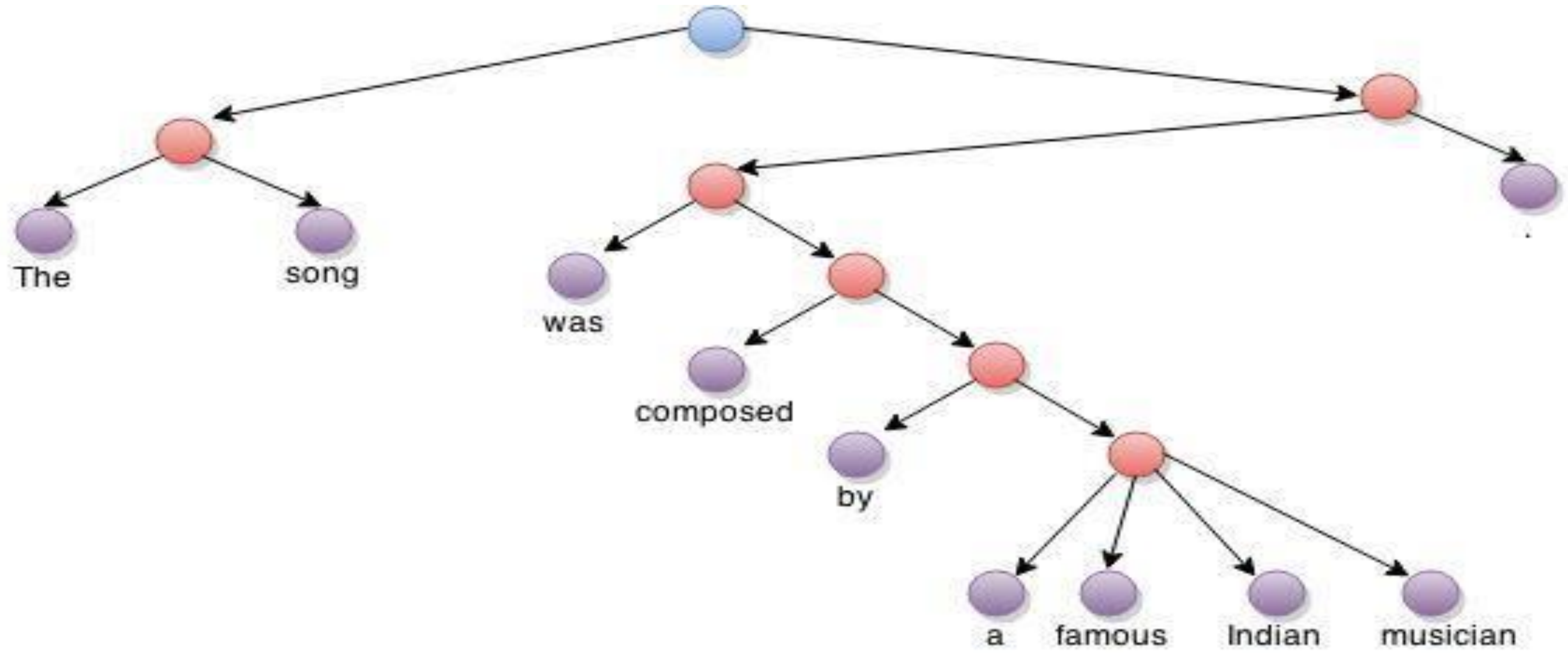


Image created using www.draw.io

Binary Parse Tree

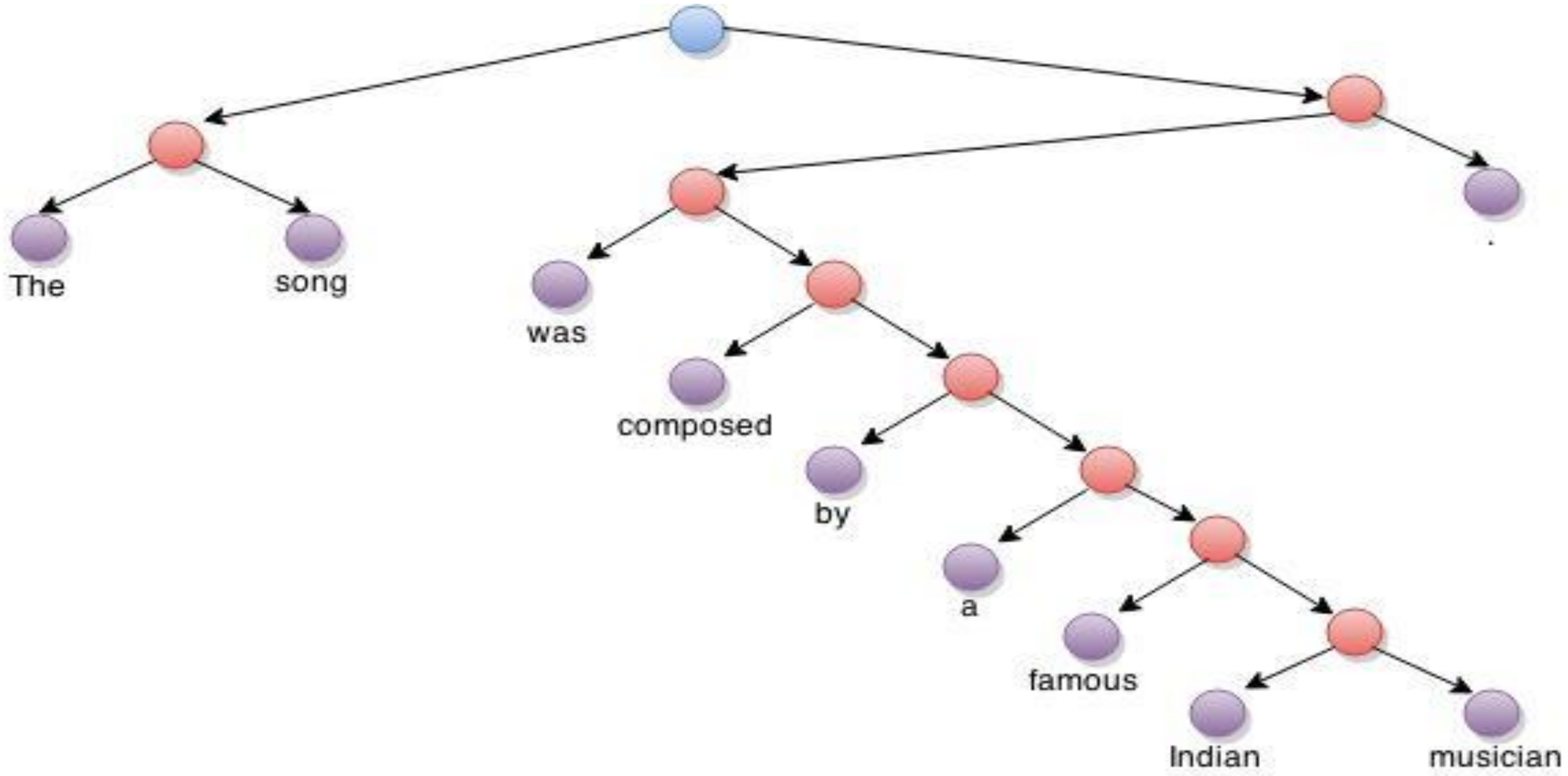


Image created using www.draw.io

What’s Novel ?

- We introduce a recursive neural network model (RNN) that learns compositional vector representations of vectors or sentences of arbitrary length or syntactic type

- We assign a vector and a matrix to every node in the parse tree
  - Vector captures the inherent meaning of the word
  - Matrix captures how the word modifies the neighboring words

- A representation for a longer phrase is computed in a bottom-up manner by recursively combining children words according to the syntactic structure in the parse tree

Recursive Matrix-Vector Model

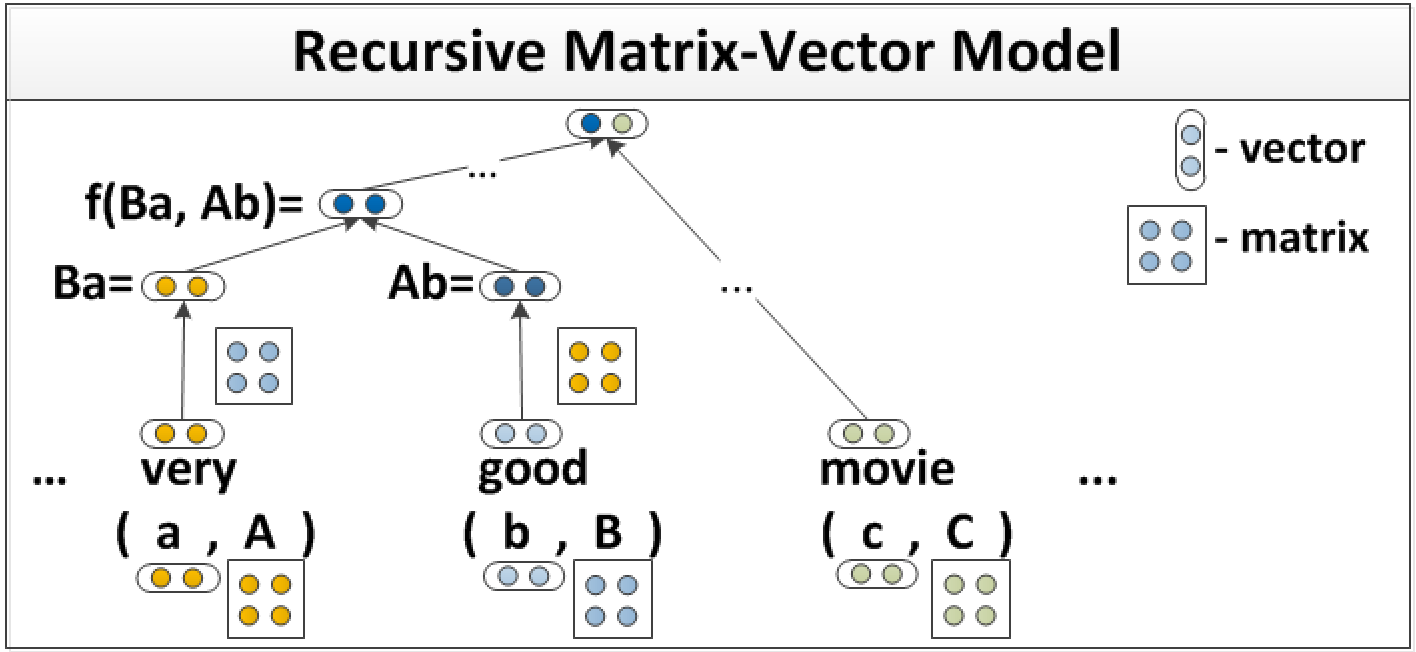


Image Source: http://www.socher.org/index.php/Main/SemanticCompositionalityThroughRecursiveMatrix-VectorSpaces

Training

- Initialize all the word vectors with pre-trained n-dimensional word-vectors
- Initialize matrices as $X = I + \varepsilon$, where I is the identity matrix and ε is Gaussian noise

- Combining two words:

$$p = f_{A,B}(a,b) = f(Ba, Ab) = g\left(W\begin{bmatrix} Ba \\ Ab \end{bmatrix}\right)$$

$$P = f_M(A,B) = W_M\begin{bmatrix} A \\ B \end{bmatrix}$$

Training

- We train vector representations by adding on top of each parent node a softmax classifier to predict a class distribution over sentiment or relationship classes

$$d(p) = softmax\left(W^{label} p\right)$$

- Error function: $E(s,t;\theta) =$ sum of cross-entropy errors at all node, where s is the sentence and t is its tree.

Learning

- Model parameters:

$$\theta = \left(W, W_M, W^{label}, L, L_M\right)$$

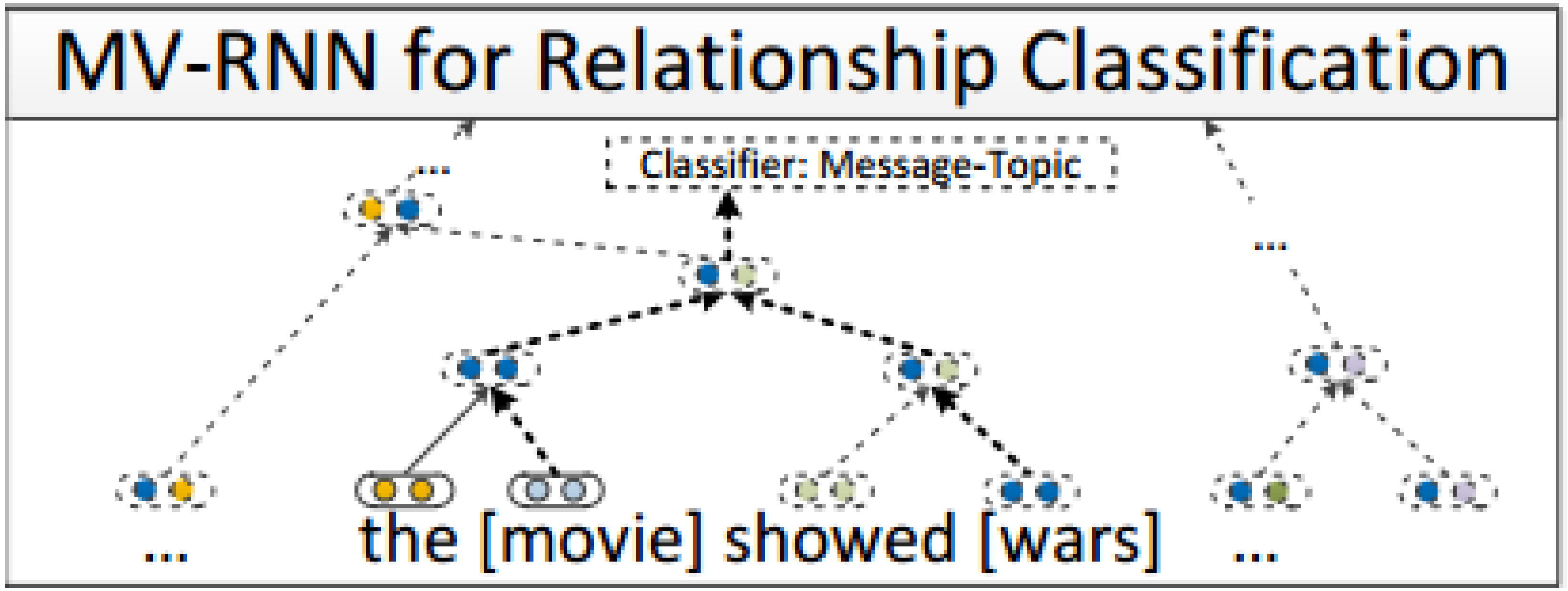
where L and L_M are the set of word vectors and word matrices.

- Objective Function:

$$\frac{\partial J}{\partial \theta} = \frac{1}{N}\sum_{(x,t)} \frac{E\left(x,t;\theta\right)}{\partial \theta} + \lambda\theta$$

where E is the cross entropy error and λ is the regularization parameter.

Classification of Semantic Relationship



Results

Dataset: SemEval 2010 Task 8

	C-E	C-W	C-C	E-D	E-O	I-A	M-C	M-T	P-P	O	<-- classified as SUM	xDIRx	ACTUAL
C-E	295	2	0	0	4	0	0	3	4	16	324	4	328
C-W	0	238	5	1	2	9	12	7	3	29	306	6	312
C-C	0	4	161	12	1	0	0	2	0	12	192	0	192
E-D	0	2	9	265	0	2	0	0	0	14	292	0	292
E-O	6	1	2	4	215	3	0	2	3	22	258	0	258
I-A	1	3	0	2	0	116	0	3	9	20	154	2	156
M-C	0	2	0	0	1	0	209	3	2	15	232	1	233
M-T	0	1	0	4	1	0	1	227	2	23	259	2	261
P-P	4	4	0	2	5	10	2	3	180	21	231	0	231
O	16	41	18	31	26	30	38	42	24	188	454	0	454
SUM	322	298	195	321	255	170	262	292	227	360	2702	15	2717

Accuracy (calculated for the above confusion matrix) = 2094/2717 = 77.07%

F1 Score = 82.51%

Code Source: http://www.socher.org/index.php/Main/SemanticCompositionalityThroughRecursiveMatrix-VectorSpaces

Reference

1. Semantic Compositionality through Recursive Matrix-Vector Spaces, Richard Socher, Brody Huval, Christopher D. Manning and Andrew Y. Ng.
   Conference on Empirical Methods in Natural Language Processing (**EMNLP 2012, Oral**)
2. Composition in distributional models of semantics, J. Mitchell and M. Lapata
   Cognitive Science,34(2010):1388–1429
3. Simple customization of recursive neural networks for semantic relation classification, Kazuma Hashimoto, Makoto Miwa, Yoshimasa Tsuruoka, and Takashi Chikayama
   2013 In EMNLP.