NLP Group : UCB

Fang Cao, Venkata Sai Mandadapu

Introduction
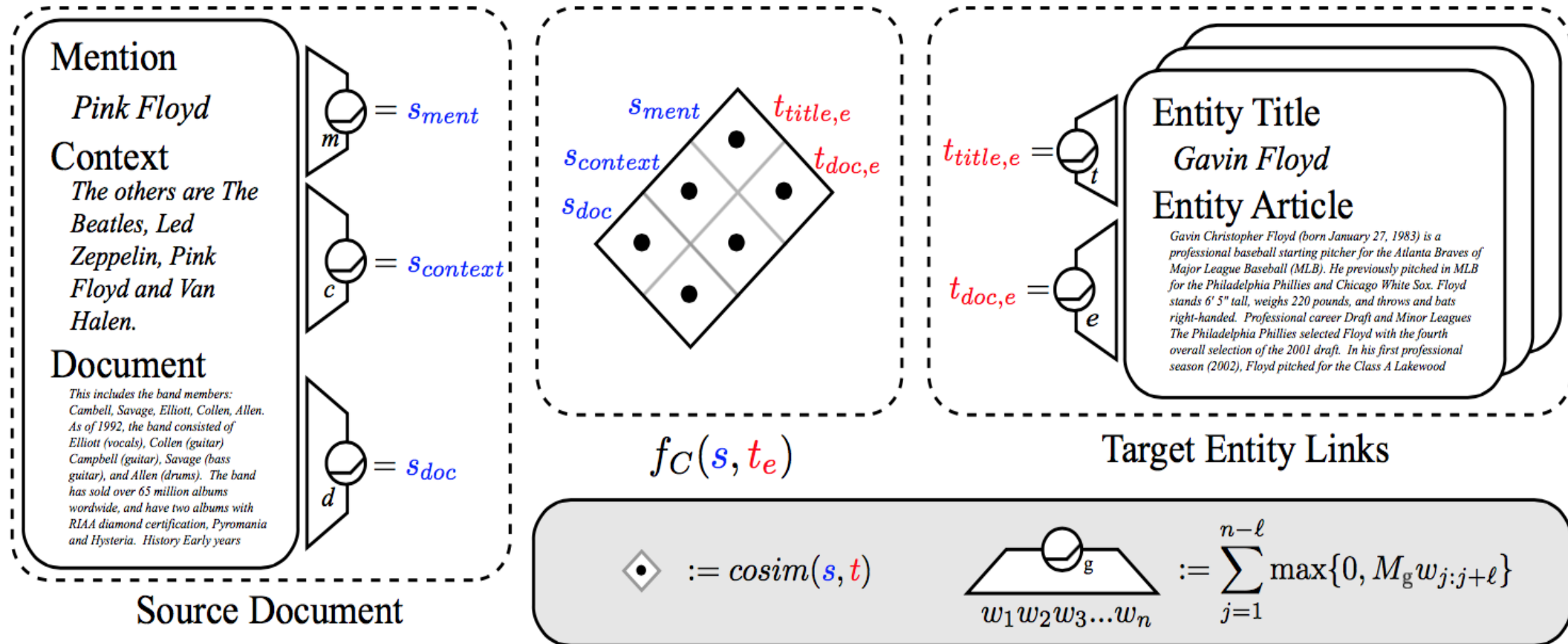
- Link to NLP group page : nlp.cs.berkeley.edu
- Faculty: Dan Klein https://people.eecs.berkeley.edu/~klein/
- Publications : http://nlp.cs.berkeley.edu/publications.shtml
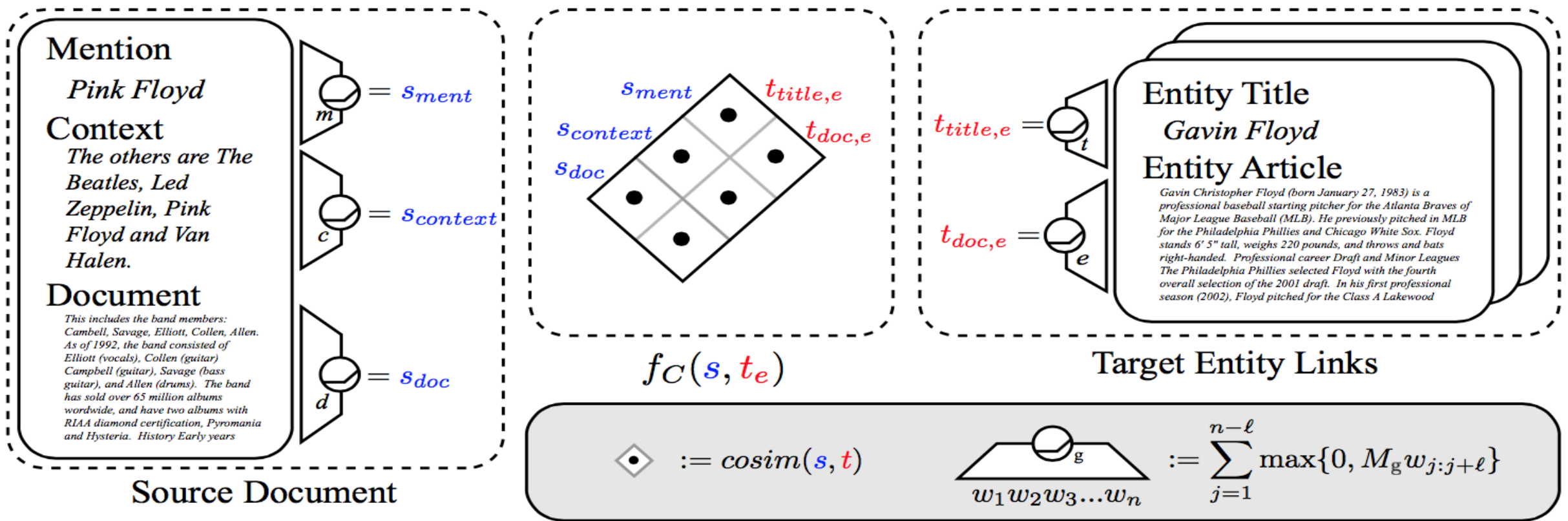- Projects: http://nlp.cs.berkeley.edu/projects/

Capturing Semantic Similarity for Entity Linking with Convolutional Neural Networks

By Matthew Francis-Landau, Greg Durrett and Dan Klein

- One of the major challenges of entity linking is resolving contextually polysemous mentions. For example, *Germany* may refer to a nation, to that nation's government, or even to a soccer team

- model semantic similarity between a mention's source document context and its potential entity targets using convolutional neural networks (CNNs)

- topic semantics at different granularities in a document are helpful in determining the genres of entities for entity linking
- CNNs can distill a block of text into a meaningful topic vector.

Source Document

$f_C(s, t_e)$

Target Entity Links

$\diamond := cosim(s, t)$

$:= \sum_{j=1}^{n-\ell} \max\{0, M_g w_{j:j+\ell}\}$

Source Document

Mention

Pink Floyd

Context

The others are The Beatles, Led Zeppelin, Pink Floyd and Van Halen.

Document

This includes the band members: Cambell, Savage, Elliott, Collen, Allen. As of 1992, the band consisted of Elliott (vocals), Collen (guitar) Campbell (guitar), Savage (bass guitar), and Allen (drums). The band has sold over 65 million albums wordwide, and have two albums with RIAA diamond certification, Pyromania and Hysteria. History Early years

$= s_{ment}$

$= s_{context}$

$= s_{doc}$

$f_C(s, t_e)$

Target Entity Links

Entity Title

Gavin Floyd

Entity Article

Gavin Christopher Floyd (born January 27, 1983) is a professional baseball starting pitcher for the Atlanta Braves of Major League Baseball (MLB). He previously pitched in MLB for the Philadelphia Phillies and Chicago White Sox. Floyd stands 6' 5" tall, weighs 220 pounds, and throws and bats right-handed. Professional career Draft and Minor Leagues The Philadelphia Phillies selected Floyd with the fourth overall selection of the 2001 draft. In his first professional season (2002), Floyd pitched for the Class A Lakewood

$t_{title,e} =$

$t_{doc,e} =$

$\diamond := cosim(s, t)$

$:= \sum_{j=1}^{n-\ell} \max\{0, M_g w_{j:j+\ell}\}$

- Embed each word into a d-dimensional vector space, yielding a sequence of vectors w1, . . . , wn. Then map those words into a fixed-size vector using a convolutional network

- fc(s, te) indicate the different types of similarity, this vector can then be combined with other sparse features and fed into a final logistic regression layer. When trained with back propagation, the convolutional networks should learn to map text into vector spaces that are informative about whether the document and entity are related or not.

Integrating with a Sparse Model

- introduce a latent variable q to capture which subset of a mention (known as a *query*) we resolve
- the set of queries defines the set of candidate entities that we consider linking a mention to
- each query generates a set of potential entities based on link counts, whose unions are then taken to give on the possible entity targets for each mention

$$P(t|x) = \sum_q P(t, q|x)$$

$$P(t, q|x) \propto \exp\left(w^\top \left(f_Q(x, q) + f_E(x, q, t) + f_C(x, t; \theta)\right)\right)$$

- f_Q and f_E are both sparse features vectors and are taken from previous work
- f_Q only impacts which query is selected and not the disambiguation to a title
- f_E mostly captures how likely the selected query is to correspond to a given entity
- Note that f_C has its own internal parameters θ because it relies on CNNs with learned filters; however, we can compute gradients for these parameters with standard back propagation. The whole model is trained to maximize the log likelihood

Single Document Summarizer

Greg Durrett, Taylor Berg-Kirkpatrick and Dan Klein

Model

- Incorporates : Compression and Anaphoricity Constraints
- Model tries to maximize number of elementary discourse units and also pronouns without losing their reference.

$$\max_{\mathbf{x}^{\text{UNIT}}, \mathbf{x}^{\text{REF}}} \left[\overbrace{\sum_i \left[x_i^{\text{UNIT}} (\mathbf{w}^\top \mathbf{f}(u_i)) \right]}^{\text{Extraction Score}} + \overbrace{\sum_{(i,j)} \left[x_{ij}^{\text{REF}} (\mathbf{w}^\top \mathbf{f}(r_{ij})) \right]}^{\text{Anaphora Score}} \right]$$

subject to

Grammaticality Constraints (Section 2.1)

$$\forall i, k \quad x_i^{\text{UNIT}} \le x_k^{\text{UNIT}} \quad \text{if } u_i \text{ requires } u_k$$

Length Constraint

$$\sum_i x_i^{\text{UNIT}} |u_i| + \overbrace{\sum_{(i,j)} x_{ij}^{\text{REF}} (|r_{ij}| - 1)}^{\substack{\text{Length adjustment} \\ \text{for explicit mention}}} \le k$$

Anaphora Constraints (Section 2.2)

$$\forall j \quad x_{ij}^{\text{REF}} = 1 \quad \text{iff no prior included textual unit mentions the entity that } r_{ij} \text{ refers to}$$

$$\forall i, k \quad x_i^{\text{UNIT}} \le x_k^{\text{UNIT}} \quad \text{if } u_i \text{ requires } u_k \text{ on the basis of pronoun anaphora}$$
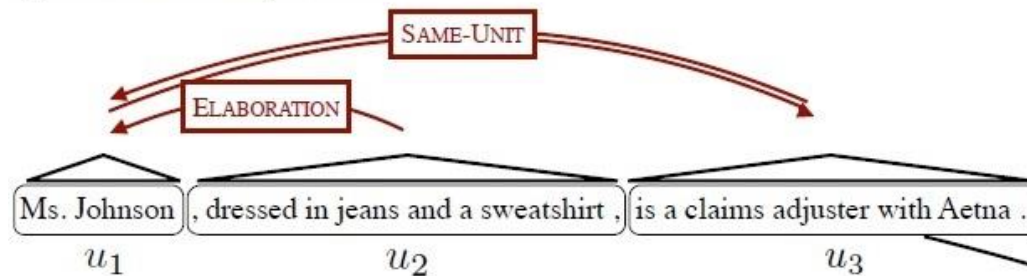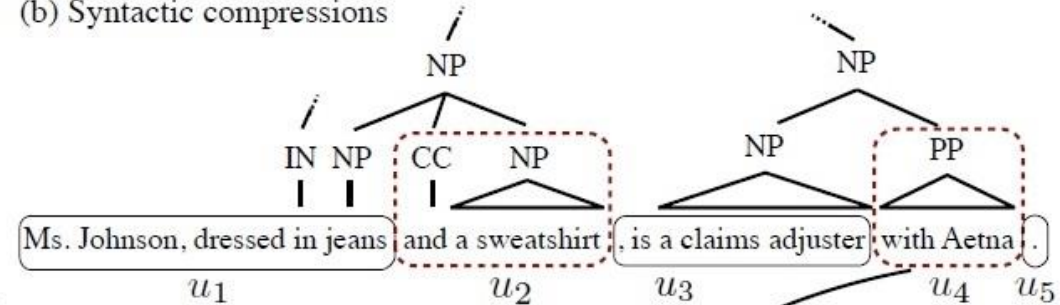
Compression

- Mixture of Rhetorical Structure Theory(RST) compression and Syntactic compression.
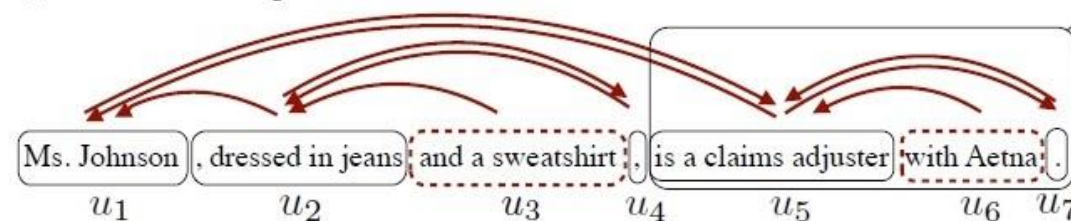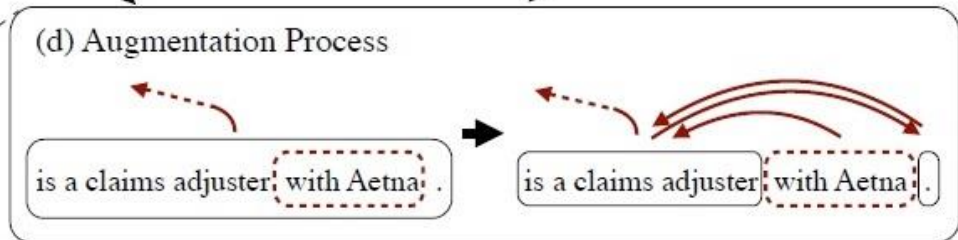

(a) Discourse compressions
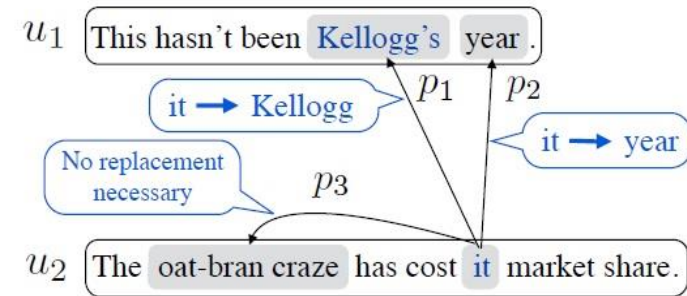
(b) Syntactic compressions

(c) Combined compressions

(d) Augmentation Process

Anaphora Constraints

- Pronoun Replacement.
- Constraining summary to include pronouns.

u_1 This hasn't been Kellogg's year .

it ➝ Kellogg

p_1 p_2

it ➝ year

No replacement necessary p_3

u_2 The oat-bran craze has cost it market share.

Replacement (2.2.1): If $\max(p_1, p_2, p_3) > \alpha$:

Allow pronoun replacement with the predicted antecedent and add the following constraint:

$x_{2,1}^{\text{REF}} = 1$ iff $x_1^{\text{UNIT}} = 0$ and $x_2^{\text{UNIT}} = 1$

Replace the first pronoun in the second textual unit

Otherwise (i.e. if no replacement is possible):

Antecedent inclusion (2.2.2): If $p_1 + p_2 > \beta$

Add the following constraint: $x_2^{\text{UNIT}} \leq x_1^{\text{UNIT}}$

Learning

- Train to maximize on features : Lexical, Structural, Centrality and Pronoun Replacement.

- Uses stochastic sub gradient descent to reduce loss, which is calculated through rogue function.

- Use extended version of the ILP model.

$$\max_{\mathbf{x}^{\text{UNIT}}, \mathbf{x}^{\text{REF}}, \mathbf{x}^{\text{NGRAM}}} \left[\sum_i \left[x_i^{\text{UNIT}} (\mathbf{w}^\top \mathbf{f}(u_i)) \right] \right.$$

$$\left. + \sum_{(i,j)} \left[x_{ij}^{\text{REF}} (\mathbf{w}^\top \mathbf{f}(r_{ij})) \right] - \ell(\mathbf{x}^{\text{NGRAM}}, \mathbf{y}) \right]$$

subject to all constraints from Figure 1

$x_i^{\text{NGRAM}} = 1$ iff an included textual unit or replacement

contains the ith reference n-gram

Summary

NYT50 article:

> Federal officials reported yesterday that students in 4th, 8th and 12th grades had scored modestly higher on an American history test than five years earlier, although more than half of high school seniors still showed poor command of basic facts like the effect of the cotton gin on the slave economy or the causes of the Korean War. Federal officials said they considered the results encouraging because at each level tested, student performance had improved since the last time the exam was administered, in 2001. "In U.S. history there were higher scores in 2006 for all three grades," said Mark Schneider, commissioner of the National Center for Education Statistics, which administers the test, at a Boston news conference that the Education Department carried by Webcast. The results were less encouraging on a national civics test, on which only fourth graders made any progress. The best results in the history test were also in fourth grade, where 70 percent of students attained the basic level of achievement or better. The test results in the two subjects are likely to be closely studied, because Congress is considering the renewal of President Bush's signature education law, the No Child Left Behind Act. A number of studies have shown that because No Child Left Behind requires states…

Summary:

National Center for Education Statistics reports students in 4th, 8th and 12th grades scored modestly higher on American history test than five years earlier. Says more than half of high school seniors still show poor command of basic facts. Only 4th graders made any progress in civics test. New exam results are another ingredient in debate over renewing Pres Bush's signature No Child Left Behind Act.

Filtered article:

> Long before President Bush's proposal to rethink Social Security became part of the national conversation, Westchester County came up with its own dialogue to bring issues of aging to the forefront. Before the White House Conference on Aging scheduled in October, the county's Office for the Aging a year ago started Speak-Up, which stands for Student Participants Embrace Aging Issues of Key Concern, to reach students in the county's 13 colleges and universities. Through a variety of events to bring together the elderly and college students, organizers said they hoped to have by this spring a series of recommendations that could be given to Washington…

Summary:

Article on Speak-Up, program begun by Westchester County Office for the Aging to bring together elderly and college students.

Thanks