An Application of Latent Semantic Analysis to Word Sense Discrimination for Words with Related and Unrelated Meanings

Juan Pino Maxine Eskenazi





Outline

Motivation

Experiment

Results

Conclusion and Future Work

Outline

Motivation

Experiment

Results

Conclusion and Future Work

Framework

- REAP: tutor for ESL vocabulary learning
- Authentic documents from the Web
- Uses cloze questions (fill-in-the-blank)

< D > < 同 > < E > < E > < E > < 0 < 0</p>

- Target words in the readings
- Questions about the target words

Lecture 14. The causes of booms and busts

from the Economist

The business cycleboom and bust. Economists still cannot agree about what causes contractions and expansions in economic activity.

Two main explanations: the business cycle as a self-perpetuating process, or shocks and policy mistakes. The main theories are:

Exogenous shocks: Recessions are caused by unexpected events, such as the rise in oil prices in the mid-1970s or, as some tried to argue, the terrorist attacks on September 11th last year. If so, recessions are, by definition, totally unpredictable. They cannot be prevented, but once they have arrived governments can use fiscal and monetary policies to cushion demand.

Keynesian theory: John Maynard Keynes blamed recessions on the inherent instability of investment caused by animal splitts: wings in the mood of producers, from optimism to pessimism. As investment slumps, Jobs and household incomes fall, amplifying the initial drop in demand. Unemployment rises because workers will not accept the pay cuts required to price the jobies back into work. So, to bring the economy back to full employment, the government needs to <u>pursue</u> expansionary policies.

Real business-cycle theory: This theory, which emerged in the early 1980s, sees productivity shocks as the cause of economic fluctuations. For example, if productivity falls, current returns decline, so workers and firms choose to work less and take more leisure. Rather than explaining the cycle in terms of market falure, as Kymes did, real business-cycle theory twes a recession as the optimal response by households and firms to a bift in productivity. If so, there is no point in government stimulating the economy. But most economists find this theory hard to swallow. Mike Mussa, a former chief economist at the IMF, describes it as the theory according to which the 1930s should be known not as the Great Depression, but the Great Vacation.

Policy mistakes: The late Rudi Dornhusch, an economist at the MIT, once remarked. None of the postwar expansions died of old age, they were all murdered by the Fed. Almost every recession size 1946, with the exception of late years', was preceded by a sharp rise in inflation that Accord central banks to raise interest rates. The first mistake was to allow economies to own-heat; the second to siam on the brakes too hard. This theory gave rise to the popular heliof that recession could be avaided so long as governments <u>pursued</u> and busits.

Austrian business-cycle theory: This is the oldest, developed by Austrian economists such as Ludwig von Mises and Friedrich Hayek in the early 20th century. Unlike Keynes, who though trecessions were caused by insufficient demand, these economists put them down to excess supply brought about by overimestment. As a result of <u>mutually</u> reinforcing movements in credit, investment and profits, each boom contains the seeds of the subsequent recession and each recession the seeds of the subsequent boom.

Output fluctuates because the short-term interest rate for loans diverges from the natural or equilibrium interest rate the rate at which the supply of saving from households equals the demand for investment fluctuates by firms. It central hanks hold interest rates below this rate, credit and investment versite most and investment versites and supply of saving from households equals creates an simulate between future output (which will increase as a result of higher investment) and future spending (which will fail as a result of lower saving today). Cheap credit and inflated profit expectations cause both overinvestment and mailwestment in the wrong kind of capital. The mismatch between saving and investment versitement will avertail by pub higher trates, making some previous investments unprofitable. Too much capacity will also reduce profits. Investment collagses, subering in a recession. As excess capacity is cut, profits rise and investment eventually recovers.

According to this theory, central banks would not be able to avoid a downturn by heading off a rise in interest rates. The only way to prevent the cycle from turning is to inject ever more credit, which becomes unsustainable. A recession is inevitable, and indeed necessary to correct the imbalance between saving and investment.

Look up a word

20 minutes left Done reading -->

◆ロト ◆母 ト ◆臣 ト ◆臣 ト ◆ 母 ト ◇ ○ ○

"Investment collapses, ushering in a recession."

<□ > < @ > < E > < E > E のQ @

Lecture 14. The causes of booms and busts

from the Economist

The business cycleboom and bust. Economists still cannot agree about what causes contractions and expansions in economic activity.

Two main explanations: the business cycle as a self-perpetuating process, or shocks and policy mistakes. The main theories are:

Exogenous shocks: Recessions are caused by unexpected events, such as the rise in oil prices in the mid-1970s or, as some tried to argue, the terrorist attacks on September 11th last year. If so, recessions are, by definition, totally unpredictable. They cannot be prevented, but once they have arrived governments can use fiscal and monetary policies to cushion demand.

Keynesian theory: John Maynard Keynes blamed recessions on the inherent instability of investment caused by animal splitts: wings in the mood of producers, from optimism to pessimism. As investment slumps, Jobs and household incomes fall, amplifying the initial drop in demand. Unemployment rises because workers will not accept the pay cuts required to price the jobies back into work. So, to bring the economy back to full employment, the government needs to <u>pursue</u> expansionary policies.

Real business-cycle theory: This theory, which emerged in the early 1980s, sees productivity shocks as the cause of economic fluctuations. For example, if productivity falls, current returns decline, so workers and firms choose to work less and take more leisure. Rather than explaining the cycle in terms of market falure, as Kymes did, real business-cycle theory twes a recession as the optimal response by households and firms to a bift in productivity. If so, there is no point in government stimulating the economy. But most economists find this theory hard to swallow. Mike Mussa, a former chief economist at the IMF, describes it as the theory according to which the 1930s should be known not as the Great Depression, but the Great Vacation.

Policy mistakes: The late Rudi Dornhusch, an economist at the MIT, once remarked. None of the postwar expansions died of old age, they were all murdered by the Fed. Almost every recession size 1946, with the exception of late years', was preceded by a sharp rise in inflation that Accord central banks to raise interest rates. The first mistake was to allow economies to own-heat; the second to siam on the brakes too hard. This theory gave rise to the popular heliof that recession could be avaided so long as governments <u>pursued</u> and busits.

Austrian business-cycle theory: This is the oldest, developed by Austrian economists such as Ludwig von Mises and Friedrich Hayek in the early 20th century. Unlike Keynes, who though trecessions were caused by insufficient demand, these economists put them down to excess supply brought about by overimestment. As a result of <u>mutually</u> reinforcing movements in credit, investment and profits, each boom contains the seeds of the subsequent recession and each recession the seeds of the subsequent boom.

Output fluctuates because the short-term interest rate for loans diverges from the natural or equilibrium interest rate the rate at which the supply of saving from households equals the demand for investment fluctuates by firms. It central hanks hold interest rates below this rate, credit and investment versite most and investment versites and supply of saving from households equals creates an simulate between future output (which will increase as a result of higher investment) and future spending (which will fail as a result of lower saving today). Cheap credit and inflated profit expectations cause both overinvestment and mailwestment in the wrong kind of capital. The mismatch between saving and investment versitement will avertail by pub higher trates, making some previous investments unprofitable. Too much capacity will also reduce profits. Investment collagses, subering in a recession. As excess capacity is cut, profits rise and investment eventually recovers.

According to this theory, central banks would not be able to avoid a downturn by heading off a rise in interest rates. The only way to prevent the cycle from turning is to inject ever more credit, which becomes unsustainable. A recession is inevitable, and indeed necessary to correct the imbalance between saving and investment.

Look up a word

20 minutes left Done reading -->

◆ロト ◆母 ト ◆臣 ト ◆臣 ト ◆ 母 ト ◇ ○ ○



"John saw the man ${\color{black} \textbf{collapse}}$ from utter exhaustion and lack of food."

Document: "Investment collapses, ushering in a recession."

Question: "John saw the man ${\bf collapse}$ from utter exhaustion and lack of food."

◆□▶ ◆□▶ ◆ □▶ ★ □▶ = □ の < @

Other possibilities

 The building could collapse if there is an earthquake. (different meaning)

Other possibilities

- The building could collapse if there is an earthquake. (different meaning)
- Social services for these very poor people will collapse if more money isn't provided by the central government. (related meaning)

▲ロ ▶ ▲周 ▶ ▲ 国 ▶ ▲ 国 ▶ ● の Q @

Other possibilities

- The building could collapse if there is an earthquake. (different meaning)
- Social services for these very poor people will collapse if more money isn't provided by the central government. (related meaning)

▲ロ ▶ ▲周 ▶ ▲ 国 ▶ ▲ 国 ▶ ● の Q @

 The economy will collapse completely if this fighting continues. (same meaning) Control how the meaning of a target word in a question relates to the meaning of this target word in a document

Control how the meaning of a target word in a question relates to the meaning of this target word in a document

Make the meanings match

- \blacktriangleright Given a document d containing the word w with meaning m
- ▶ Given n questions q₁, ..., q_n about w with meanings m₁, ..., m_n

◆□▶ ◆□▶ ◆ □▶ ★ □▶ = □ の < @

Find the questions q_i such that m_i = m

- \blacktriangleright Given a document d containing the word w with meaning m
- ▶ Given n questions q₁, ..., q_n about w with meanings m₁, ..., m_n

- Find the questions q_i such that m_i = m
- Simpler: find at least one question q_i such that m_i = m

Algorithms

Lesk

- Latent Semantic Analysis (LSA)
- Suitable for discrimination rather than disambiguation

◆□▶ ◆□▶ ◆ □▶ ★ □▶ = □ の < @

Lesk (one variant)

- ▶ Word **w** in sentence **s**.
- **w** has definitions $\mathbf{d}_1, \dots, \mathbf{d}_n$ in a dictionary.
- Select context **c** for **w** (e.g. **s** or words around **w**).
- Select the definition that has the most words in common with c.

Lesk for this task

- ▶ Word **w** in document **d**.
- w has questions $\mathbf{q}_1, ..., \mathbf{q}_n$.
- Select contexts \mathbf{c}_d and \mathbf{c}_q for \mathbf{w} in \mathbf{d} and in \mathbf{q} .
- Select the question that has the most words in common with c.

◆□▶ ◆□▶ ◆三▶ ◆三▶ ○三 のへぐ

Latent Semantic Analysis (LSA)

- Challenge: the questions are a short context
- LSA: initially LSI, used to match (short) queries to documents

◆□▶ ◆□▶ ◆三▶ ◆三▶ ○三 のへぐ

LSA Description

- X: term-document matrix (e.g. X_{i,j} = 1 if word i appears in document j)
- $X = U \cdot S \cdot V^T$ (SVD theorem)

•
$$U^T \cdot U = I_m$$
 and $V^T \cdot V = I_n$

- ▶ U is truncated, replaced by U_r , V replaced by V_r , S replaced by S_r
- $\blacktriangleright X \simeq U_r \cdot S_r \cdot V_r^T$
- Similarity between **d** and **q**: compute cosine similarity between $S_r^{-1} \cdot U_r^T \cdot \mathbf{d}$ and $S_r^{-1} \cdot U_r^T \cdot \mathbf{q}$

ション ヘロ・ イロ・ イロ・ イロ・

Motivation

Cope with sparsity

◆□ > ◆□ > ◆ Ξ > ◆ Ξ > → Ξ = ∽ ۹ < ↔

Remove noise

Motivation

- Cope with sparsity
 - Compute similarity in another space

◆□▶ ◆□▶ ◆ □▶ ★ □▶ = □ の < @

- Remove noise
 - truncated SVD

Experimental Goal

 Compare LSA performance for words with related and unrelated meanings to a random baseline and Lesk

◆□▶ ◆□▶ ◆三▶ ◆三▶ →□ ● ◇◇◇

- Example (related meaning): comprise
- Example (unrelated meaning): bank

Outline

Motivation

Experiment

Results

Conclusion and Future Work

Setup

- 62 manually generated cloze questions
- 16 target words (e.g. bond, collapse, compound, comprise, etc.)

< D > < 同 > < E > < E > < E > < 0 < 0</p>

- Senses manually annotated (WordNet)
- ▶ Web documents containing target word/sense gathered

Factors Studied

 Words with related and unrelated meanings vs. with unrelated meanings only

- Truncation parameter
- Context selection

Outline

Motivation

Experiment

Results

Conclusion and Future Work

Baseline

- Random Baseline: 44%
- Lesk Algorithm: 61%
- Lesk preprocessing
 - Iower casing
 - stemming
 - stop words and punctuation discarded
 - window: 10 words around the target in the question
 - window: 2 sentences around the target in the document

▲ロ ▶ ▲周 ▶ ▲ 国 ▶ ▲ 国 ▶ ● の Q @

LSA



Accuracy vs. Dimension Reduction for Related Meanings with Different Contexts

- words with related and unrelated meanings
- LSA better only with certain truncation parameter values

LSA



Accuracy vs. Dimension Reduction for Unrelated Meanings with Different Contexts

- words with unrelated meanings
- context selection plays important role

Outline

Motivation

Experiment

Results

Conclusion and Future Work

Conclusion and Future Work

- Comparison performance of Lesk, LSA and a random baseline
- Factors: truncation parameter, context selection and type (unrelated/related vs. unrelated)
- Future work: Comparison of LSA with other second-order representations of documents and questions

◆□▶ ◆□▶ ◆□▶ ◆□▶ ▲□ ◆ ○ ◆

Future work: use WSD to rank definitions

۷	REAP Rea	ading - Mozilla Firefox	a x
+ http://reap-2.cs.cmu.edu/summer09/present	group_reading.php?usemame=jmpino4&sessio	onid=64&schema=reap_summer09&condition=0&calloutwords=&calloutpos=&calloutroots=&document_id=43929	1
governments can use part Dictiona	ry Search - MozillaFirefox 📃 🗆 🕻	nt instability of investment caused by animal spirits, swings in the mood of	^
producers, from opt	s.cmu.edu/summer09/word_lookup.phtml?f	nd household incomes fall, amplifying the initial drop in demand. Unemployment rises	3
because workers will r needs to pursue expa		ack into work. So, to bring the economy back to full employment, the government	
Real business-cycle • when a pers	on or structure becomes too weak to	Os, sees productivity shocks as the cause of economic fluctuations. For example, if	
failure, as Keynes did, point in governments	failure of a system, organization, c	vork less and take more leisure. Rather than explaining the cycle in terms of market ptimal response by households and firms to a shift in productivity. If so, there is no his theory hard to evalue will be Musca a former shift opportunity at the IMM described	
it as the theory accord collapse (v)		reat Depression, but the Great Vacation.	
Policy mistakes: The the Fed Almost every • to fall down	suddenly because of pressure or	remarked: None of the postwar expansions died of old age, they were all murdered by a sharp rise in inflation that forced central banks to raise interest	2
rates. The first mistak If someone	of for support collapses they fall down because they	slam on the brakes too hard. This theory gave rise to the popular belief that	
recessions could be av are ill or we	ak	etary policies to keep inflation low and stable. Yet the recessions in Japan after the	
1980S Dubble and Ame so it can be	stored, or (especially of furniture) to	it booms and busts.	
Austrian business-c fold in this v	vay	n economists such as Ludwig von Mises and Friedrich Hayek in the early 20th century	6
overinvestment. As a inability to c	continue or work correctly	, these economists put them down to excess supply brought about by investment and profits, each boom contains the seeds of the subsequent recession	
and each recession the		,	
Output fluctuates bec		rom the natural or equilibrium interest rate the rate at which the supply of saving	
from households equa • He was take	n to hospital after his collapse on	al banks hold interest rates below this rate, credit and investment will rise too	
rapidly, and consumer the pitch.	of a tower block during the	veen future output (which will increase as a result of higher investment) and future	
spending (which will is malinyestment in the earthquake	2 MT 2003 00 2 MAN	and inflated profit expectations cause both overinvestment and	
investments unprofita	what caused the collapse [of] her	tment collapses, ushering in a recession. As excess capacity is cut, profits rise and	
investment eventually • A poor econ	omy has caused the collapse [of]	The second se	
According to this theo	f small businesses.	urn by heading off a rise in interest rates. The only way to prevent the cycle from	
turning is to inject event the brink/ve	s between the two countries are ion rge off collapse (very soon going to	ession is inevitable, and indeed necessary to correct the imbalance between saving	
and investment. fail).	Sector and a sector secto	×	
			- C
Look up a word		18 minutos left Done reading>	1

Done

N/A

Thank you. Questions?

Words and questions with annotated senses available at www.cs.cmu.edu/ $\sim\!jmpino/questions.xls$

・ロト・4回ト・ミミト・ミニー つへで