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)
- Target words in the readings
- Questions about the target words
Lecture 14. The causes of booms and busts

from the Economist

The business cycle boom 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 spirits: swings 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 jobless back into work. So, to bring the economy back to full employment, the government needs to pursue 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 failure, as Keynes did, real business-cycle theory views a recession as the optimal response by households and firms to a shift in productivity. If so, there is no place in governments 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 Dornbusch, 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 since 1945, with the exception of last year's, was preceded by a sharp rise in inflation that forced central banks to raise interest rates. The first mistake was to allow economies to overheat, the second to slam on the brakes too hard. This theory gave rise to the popular belief that recessions could be avoided so long as governments pursued prudent monetary policies to keep inflation low and stable. Yet the recessions in Japan after the 1980s bubble and America now suggest that price stability does not prevent booms and busts.

**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 thought recessions were caused by insufficient demand, these economists put them down to excess supply brought about by overinvestment. As a result of mutually 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 at which the supply of saving from households equals the demand for investment funds by firms. If central banks hold interest rates below this rate, credit and investment will rise too rapidly, and consumers will not save enough. This creates a mismatch between future output (which will increase as a result of higher investment) and future spending (which will fall as a result of lower saving today). Cheap credit and inflated profit expectations cause both overinvestment and malinvestment in the wrong kind of capital. The mismatch between saving and investment will eventually push up interest rates, making some previous investments unprofitable. Too much capacity will also reduce profits. Investment collapses, ushering 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.
“Investment collapses, ushering in a recession.”
Lecture 14. The causes of booms and busts
from the Economist

The business cycle boom 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**: Recession 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 spirits: swings 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 jobless back into work. So, to bring the economy back to full employment, the government needs to pursue 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 failure, as Keynes did, real business-cycle theory views a recession as the optimal response by households and firms to a shift in productivity. If so, there is no point in governments 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 Rudiger Dornbusch, 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 since 1945, with the exception of last year's, was preceded by a sharp rise in inflation that forced central banks to raise interest rates. The first mistake was to allow economies to overheat, the second to slam on the brakes too hard. This theory gave rise to the popular belief that recessions could be avoided as long as governments pursued prudent monetary policies to keep inflation low and stable. Yet the recessions in Japan after the 1980s bubble and America now suggest that price stability does not prevent booms and busts.

**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 thought recessions were caused by insufficient demand, these economists put them down to excess supply brought about by overinvestment. As a result of mutually 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 at which the supply of saving from households equals the demand for investment funds by firms. If central banks hold interest rates below this rate, credit and investment will rise too rapidly, and consumers will not save enough. This creates a mismatch between future output (which will increase as a result of higher investment) and future spending (which will fall as a result of lower saving today). Cheap credit and inflated profit expectations cause both overinvestment and misinvestment in the wrong kind of capital. The mismatch between saving and investment will eventually push up interest rates, making some previous investments unprofitable. Too much capacity will also reduce profits. Investment collapses, ushering 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.
John saw the man ___ from utter exhaustion and lack of food.

- collapse
- occur
- priority
- reverse
- trace

Done

“John saw the man **collapse** from utter exhaustion and lack of food.”
Document: “Investment collapses, ushering in a recession.”

Question: “John saw the man collapse from utter exhaustion and lack of food.”
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)
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)
- The economy will **collapse** completely if this fighting continues. (same meaning)
Goal

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
Given a document $d$ containing the word $w$ with meaning $m$

Given $n$ questions $q_1, \ldots, q_n$ about $w$ with meanings $m_1, \ldots, m_n$

Find the questions $q_i$ such that $m_i = m$
Given a document $d$ containing the word $w$ with meaning $m$

Given $n$ questions $q_1, \ldots, q_n$ about $w$ with meanings $m_1, \ldots, m_n$

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
Lesk (one variant)

- Word $w$ in sentence $s$.
- $w$ has definitions $d_1, ..., d_n$ 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 $q_1$, $q_2$, $q_n$.
- Select contexts $c_d$ and $c_q$ for $w$ in $d$ and in $q$.
- Select the question that has the most words in common with $c$. 
Latent Semantic Analysis (LSA)

- Challenge: the questions are a short context
- LSA: initially LSI, used to match (short) queries to documents
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 $S_r$
- $X \simeq U_r \cdot S_r \cdot V_r^T$
- Similarity between $d$ and $q$: compute cosine similarity between $S_r^{-1} \cdot U_r^T \cdot d$ and $S_r^{-1} \cdot U_r^T \cdot q$
Motivation

- Cope with sparsity
- Remove noise
Motivation

- Cope with sparsity
  - Compute similarity in another space
- Remove noise
  - truncated SVD
Experimental Goal

- Compare LSA performance for words with related and unrelated meanings to a random baseline and Lesk
  - Example (related meaning): comprise
  - Example (unrelated meaning): bank
Setup

- 62 manually generated cloze questions
- 16 target words (e.g. bond, collapse, compound, comprise, etc.)
- 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
  - lower 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
- words with related and unrelated meanings
- LSA better only with certain truncation parameter values
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
Keynesian theory: Just as producers, from profits, start to hire workers because workers will not only consume the goods they produce, but needs to pursue expansion.

Real business-cycle theory: Productivity falls, current account deficit, and then the government needs to pursue expansion.


Austrian business-cycle theory: Unlike Keynesians, who think of recessions as caused by overinvestment. As output grows and each recession the growth in output fluctuates between investment and consumption, but also there is malinvestment in the sense that investment is unprofitable and then investment eventually falls.

Output fluctuates because of shocks to households equality, rapid consumption in the 1960s, and consumer spending (which will fall), and some expenditures are a malinvestment in that investments that are profitable in the long run will be unprofitable in the short run.

According to this theory, the turning point is to inject even more investment and consumption.

**Examples:**

- He was taken to hospital after his collapse on the pitch.
- The collapse of a tower block during the earthquake.
- I don't know what caused the collapse of her marriage.
- A poor country has caused the collapse of thousands of small businesses.
- Negotiations between the two countries are on the brink of collapse (very soon going to fail).

**Instability of investment caused by animal spirits:** Swings in the mood of businesspersons, household incomes fall, amplifying the initial drop in demand. Unemployment rises back into work. So, to bring the economy back to full employment, the government can adopt a policy of greater deficit spending, which will increase the output, and reduce the unemployment.

0s, see productivity shocks as the cause of economic fluctuations. For example, if work less and take more leisure. Rather than explaining the cycle in terms of market optimality response by households and firms to a shift in productivity. If so, there is no reason to worry about the Great Depression, but the Great Vacation.

remarked: None of the postwar expansions died of old age, they were all murdered by the Fed. Was preceded by a sharp rise in inflation that forced central banks to raise interest rates, which brought the boom to a halt.

In economists such as Ludwig von Mises and Friedrich Hayek, in the early 20th century, recessions are caused by a lack of investment, which leads to a decrease in output and profits. Consequently, this induces a depression, which eventually leads to a recession. As excess capacity is cut, profits rise and output increases, leading to a recovery.

From the natural or equilibrium interest rate, the output, and inflation, profit expectations cause both overinvestment and underinvestment. If saving and investment will eventually push up interest rates, making some previous investments unprofitable and future investment unprofitable. As excess capacity is cut, profits rise and output increases, leading to a recovery. The only way to prevent the cycle from occurring is inevitable, and indeed necessary to correct the imbalance between saving and investment.
Thank you. Questions?

Words and questions with annotated senses available at www.cs.cmu.edu/~jmpino/questions.xls