

# Lecture 16: Implicit Relevance Feedback and Syntactic Structure

Scribes: Ellis Weng, Andrew Owens

April 13, 2010

## 1 Implicit Relevance Feedback

In the previous lecture, we introduced *implicit* relevance feedback. The goal of implicit relevance feedback is to infer which search results are relevant from the user's normal interactions with the system. One example of implicit relevance feedback is clickthrough data – information about which search results the user clicks on.

Clickthroughs only give *relative* relevance judgments. The result is a *partial* ordering of some documents with respect to a query. For example, for documents  $d_3$ ,  $d_5$ , and  $d_{14}$  we may have

$$\begin{aligned}\text{rel}_q(d_{14}) &> \text{rel}_q(d_3) \\ \text{rel}_q(d_{14}) &> \text{rel}_q(d_5).\end{aligned}$$

In this example, we know  $d_{14}$  is more relevant than  $d_3$  and  $d_5$ , but we know nothing about the relevance of  $d_3$  versus that of  $d_5$ .

In the previous lecture, we introduced an algorithm for ranking documents with respect to these relative constraints [1]. The idea is to rank documents by the magnitude of their orthogonal projections onto a vector  $\vec{w}$ . Documents with greater projections rank higher. As far as possible, this ranking should respect the partial ordering we obtained from clickthrough data.

This problem as stated is underconstrained, since there may be many  $\vec{w}$  vectors that preserve the partial ordering equally well. Also, there may not exist a  $\vec{w}$  that satisfies all the constraints obtained. To break ties, we add the additional constraint that  $\vec{w}$  maximizes the distance between the closest pair of projections.

### 1.1 Generalization

The  $\vec{w}$  that we have chosen above is query-specific. However, it would be convenient to have a *single*  $\vec{w}$  that would work across multiple queries. One way to do this is to encode information about  $q$  and  $d$ 's "compatibility" in the

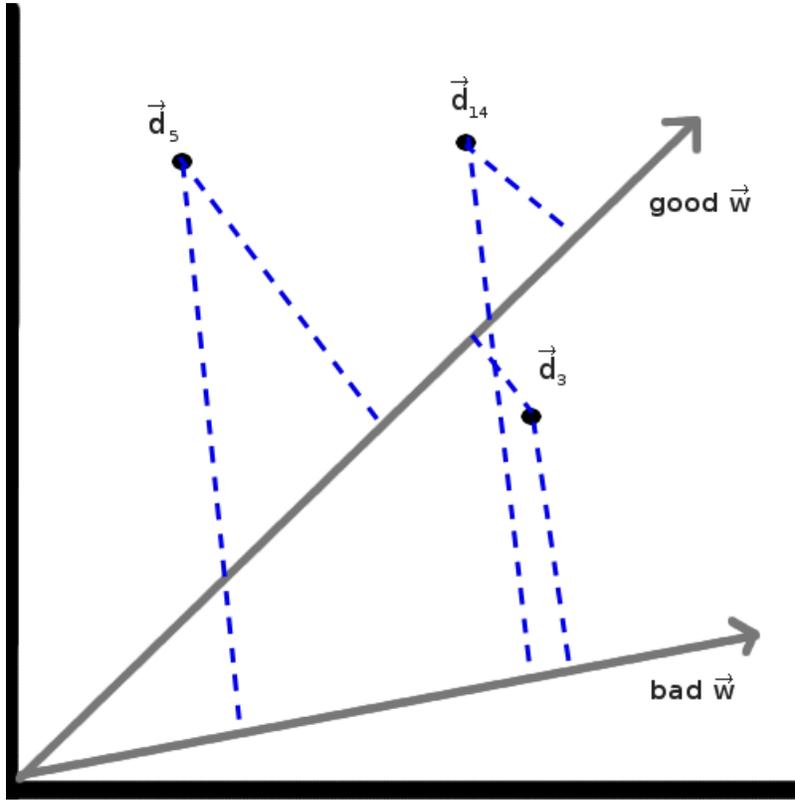


Figure 1: This figure shows the document vectors  $\vec{d}_3$ ,  $\vec{d}_5$ , and  $\vec{d}_{14}$  projected onto two  $\vec{w}$  vectors. With the “good”  $\vec{w}$ , the documents respect the partial order relation defined above. The “bad”  $\vec{w}$  violates the constraint  $\text{rel}_q(d_{14}) > \text{rel}_q(d_3)$ .

vector representation of  $d$ . We replace  $\vec{d}$  with the new vector  $\vec{\phi}_{q,d}$ . Here are some possible entries of  $\vec{\phi}_{q,d}$ :

$$\vec{\phi}_{q,d}[1] = \cos(\angle(q, d))$$

where  $\angle(q, d)$  is the angle between  $\vec{q}$  and  $\vec{d}$ ,

$$\vec{\phi}_{q,d}[2] = \begin{cases} 1 & \text{if } d \text{ contains a word in } d' \text{'s URL domain name} \\ 0 & \text{otherwise} \end{cases}$$

$$\vec{\phi}_{q,d}[3] = \begin{cases} 1 & \text{if Google, Yahoo!, or Bing puts } d \text{ in their top 10 search results} \\ 0 & \text{otherwise} \end{cases}$$

This last feature is an example of a “meta search feature.” With it, our search engine is essentially using the results of other search engines to improve its results! Note that features need not be binary (as already shown by  $\phi_{q,d}[1]$ ). For

example, we could introduce a feature that is the document's rank on Google. Other possible features include the following: an indicator that says whether  $d$  is the title of a Wikipedia document, a field denoting the country where  $d$  was written, or an indicator that says whether  $d$ 's language matches that of  $q$ .

Note that the model allows for different users to have different feature vectors and thus different  $\vec{w}$  vectors.

## 2 Syntactic structure

The understanding of syntactic structure plays a role in many applications, especially in the field of information extraction. The document retrieval paradigms that we have discussed so far largely ignore syntactic structure. For more complicated tasks (like automatically extracting facts and opinions), we (seemingly) need a deeper understanding of sentence structure. Machine translation, sentence compression, question answering and database population are all problems that require us to understand sentence structure.

Why is structure necessary for these applications? From an Information Retrieval perspective, structure has generally been completely ignored; queries and documents were essentially treated as a bag of words. For example, the sentences "Microsoft bought Google" and "Google bought Microsoft" would be indistinguishable in these models. Clearly, these two sentences have different meanings, even though they have the same set of words. In order to understand the meanings of sentences, we have to rely on sentence structure, which is not a simple task. One might argue that this sentence is easy to understand because the company after "bought" is the entity that is being bought. However, there are many exceptions to such simple rules, for example, "Microsoft was bought by Google." In this example, we see that Microsoft is the company being bought even though Google appears after the word "bought." This example shows the difficulty in understanding the meaning of a sentence. We will run into many more complex problems when analyzing syntactic structure.

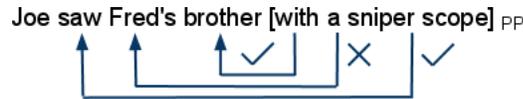
## 3 Ambiguity

One common problem that we have to deal with when analyzing syntactic structure is ambiguity. There are several types of ambiguity, such as:

1. prepositional phrase attachment ambiguity
2. part-of-speech ambiguity
3. word sense ambiguity

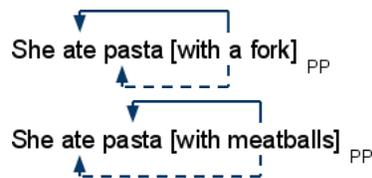
We will now cover each of these errors in depth.

### 1. Prepositional Phrase Attachment Ambiguity



In this example, the arrows represent who has the telescope; the arrows do not indicate where the prepositional phrase attaches to. This is an example of prepositional phrase attachment ambiguity; it is unclear what the prepositional phrase “with a sniper scope” attaches to. Did Joe use a sniper scope to see Fred’s brother or did Fred’s brother have a sniper scope? This prepositional phrase can either be attached to the verb “saw” or the noun phrase “Fred’s brother.” Either interpretation is valid; however, note that the set of who can own the scope is not the set of all people mentioned in the sentence. For example, in this sentence Fred cannot be the owner. There is some underlying structure that needs to be explored in order to understand why this is the case.

In addition to the actual structure of the sentence, the lexical items (or vocabulary) also play an important role in deciphering this ambiguity. Here are two sentences that are relatively non-ambiguous:



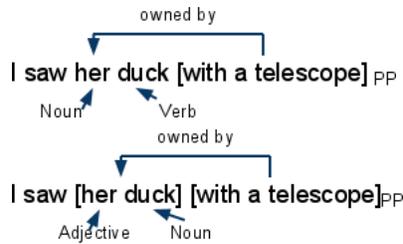
Although these two sentences could potentially have the same structure ambiguity as the first example, the possible intended meaning of each sentence is fairly clear. The solid lines show what the prepositional phrases are modifying, and the dotted lines show a potential alternative meaning. These sentences are not nearly as ambiguous as the first sentence because of the lexical items. This example shows how the sentence structure is not the only thing that determines the meaning of a sentence, since the lexical items play a large role as well.

## 2. Part-of-Speech Ambiguity

Another type of ambiguity is part-of-speech ambiguity. Here is an example that is rather famous for being very ambiguous:

I saw her duck [with a telescope] PP

First, note that there is already PP attachment ambiguity, where “with a telescope” can modify either “saw” or “her duck.” The words “her duck” have the second type of ambiguity. “Her” can either be an adjective or noun, and correspondingly, “duck” can be either a noun or verb. This part-of-speech ambiguity creates two different interpretations for the same sentence:



The prepositional phrase can now relate to either "her" in the case where "her" is a noun and "duck" is a verb, or it can relate to "her duck" in the case where "her" is a pronoun and "duck" is a noun.

### 3. Word Sense ambiguity

There is also word sense ambiguity in the following sentence:

I saw her duck [with a telescope] PP

The word "saw" has multiple meanings. This is different from a part-of-speech ambiguity because the different senses of the word "saw" are all verbs. Each of these different senses can lead to at least one more interpretation of the sentence. For example, the word "saw" can be interpreted as the perceiving through the eyes "saw," which is probably the original meaning of the sentence. However, it can also mean the action of "sawing" or even in a poker sense of matching a bet.

Although this sentence has around 10 different (technically) valid interpretations, there are only a few sentences that make sense to a normal person. For example, this sentence could either mean that a person looked through a telescope and saw a female in the action of ducking, or it could mean that a person used a telescope as a cutting instrument to saw his female friend's duck. One situation is probably more likely to occur than another based on previous experiences in life and common knowledge. In probabilistic terms, each person assigns a prior probability to each sentence occurring. These prior probabilities help determine which interpretations the reader finds plausible.

## 4 Machine Translation

The syntactic structure of sentences is particularly important for machine translation, as the structure of a sentence varies with the language. For example, English sentences typically appear in a subject, verb, object (SVO) order, whereas Hindi appears in subject, object, verb (SOV) order, and Irish appears in verb, subject, object (VSO) order. Each language has a specific order that meaningful sentences must generally conform to. Thus, it is important to understand the structural divergences between languages in order to translate from one language to another.

## 5 Sentence Compression (Summarization)

Sentence compression refers to the shortening of sentences to get to the core meaning of a sentence. Here is an example:

Betta put (her signature ginger) cookies [on the table][with a flourish].

The diagram shows the sentence "Betta put (her signature ginger) cookies [on the table][with a flourish]." with blue arrows indicating parameter status. A top arrow labeled "optional" spans from "(her signature ginger)" to "[with a flourish]". A bottom arrow labeled "non-optional" spans from "(her signature)" to "(ginger)".

This sentence could be compressed by removing the optional parameters.

**Betta put cookies [on the table]**

However, if one or more of the non-optional parameters are removed, the sentence is no longer legal.

**\*Betta put cookies [with a flourish]**

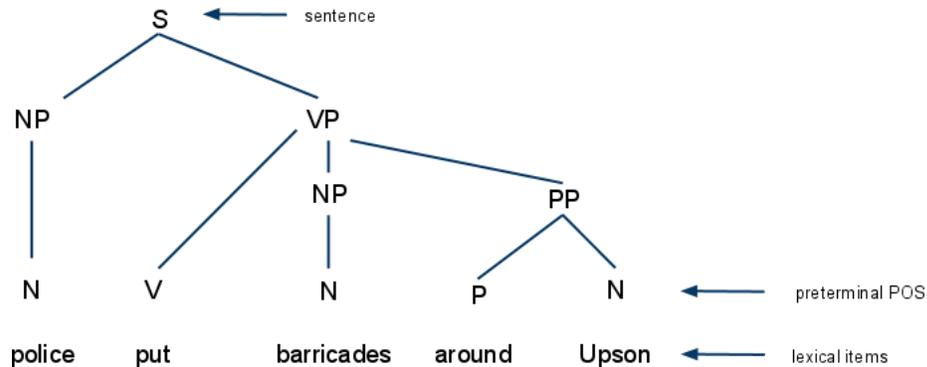
(The star in front of a sentence denotes a syntactically incorrect sentence.)

This sentence is invalid because the prepositional phrase “with a flourish” is the only argument to the action verb “put”, which is invalid because there needs to be a location somewhere after the noun “cookies” because “put” requires two parameters: a noun phrase (the object that is being moved) and a prepositional phrase (the location the object is being moved to).

Although “with a flourish” is an argument of “put”, there is still a syntactic error because a location was not specified. It could be rather difficult to see an error in this form because the structure of these sentences look similar. Both these sentences have the following sequence: Noun, verb, noun, prepositional phrase. Yet, one is grammatically correct, and the other one is not. There are many lexical considerations that are involved in determining the legality of a sentence.

## 6 Constituent Analysis

Constituent analysis is one class of formalisms for describing legal structures. Constituents can be thought of as hierarchical multi-word types. Here is an example of constituent analysis for the sentence “Police put barricades around Upson”:



This structure clearly shows how the lexical items are grouped into constituents that are related to each other. This structure exhibits “XP regularity”, which means that each XP has an X as the head. The controlling lexical item is the head of the XP. For example, the head of the VP “put barricades around Upson” is the verb “put.”

## 7 Context-Free Grammar (CFG)

A common formalism for describing licit constituent structures is a context-free grammar. There are several key advantages to CFGs. First, CFGs are extremely simple and flexible. Second, structures can be constructed efficiently; there are efficient dynamic programming algorithms that can be used to recover the structure of a sentence. In the next lecture, we will further discuss these advantages and some disadvantages of Context-Free Grammars.

## 8 Questions

- Suppose we have three documents whose feature vectors for a particular query  $q$  are:  $\vec{\phi}_{q,d_1} = [1, 1]^T$ ,  $\vec{\phi}_{q,d_2} = [0, 1]^T$ , and  $\vec{\phi}_{q,d_3} = [1, 0]^T$ . Furthermore, assume our clickthrough data gives a partial ordering on the relevance:  $\text{rel}(d_1) > \text{rel}(d_2)$  and  $\text{rel}(d_1) > \text{rel}(d_3)$ .
  - Find a vector  $\vec{w}$  that preserves this partial ordering when the documents are ranked by their orthogonal projection onto  $\vec{w}$ .
  - Suppose that we introduce  $d_4$  such that  $\text{rel}(d_4) > \text{rel}(d_1)$  and  $\vec{\phi}_{q,d_4} = [-1, 0]^T$ . Show that there is no  $\vec{w}$  that satisfies this partial ordering. How could we fix this problem?
- Because headlines need to be short and to the point, important words are often eliminated, optional or not. As a result, many headlines can be interpreted in different ways (and can often be humorous). What is the

ambiguity in the following sentences? If there is a prepositional phrase attachment ambiguity or part-of-speech ambiguity, draw a constituent tree for each different interpretation. The following sentences come from [2].

- (a) Drunk gets nine months in violin case.
  - (b) Stolen painting found by tree.
  - (c) Eye drops off shelf.
  - (d) Actor sent to jail for not finishing sentence.
  - (e) Man shoots neighbor with machete.
  - (f) Chou remains cremated.
3. (a) The problem of automatically generating a headline from a newspaper article can be considered a sentence compression problem. How would you modify a “standard” English CFG for headline-style compression?
- (b) What are some structural and semantic criteria you would use to analyze the quality of a compressed sentence? Describe some automated techniques for evaluating sentences based on these criteria.

## 9 Solutions

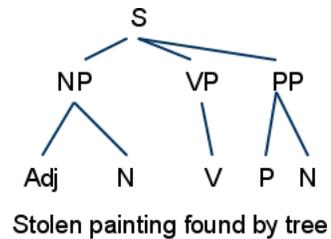
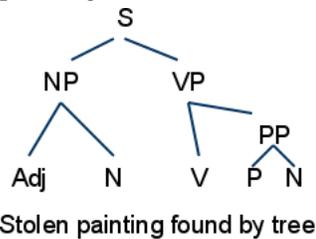
1. (a) Let  $w = [w_1, w_2]^T$ . We require  $\vec{w} \cdot \vec{\phi}_{q,d_1} > w \cdot \vec{\phi}_{q,d_2}$  and  $\vec{w} \cdot \vec{\phi}_{q,d_1} > \vec{w} \cdot \vec{\phi}_{q,d_3}$ . The first case is equivalent to  $w_1 + w_2 > w_2$  and the second case is equivalent to  $w_1 + w_2 > w_1$ . Thus any positive choice of  $w_1$  and  $w_2$  satisfies the partial ordering.
- (b) Suppose by way of contradiction that there is such a  $\vec{w}$ . Then  $\vec{w} \cdot \vec{\phi}_{q,d_1} + \vec{w} \cdot \vec{\phi}_{q,d_1} + \vec{w} \cdot \vec{\phi}_{q,d_4} > \vec{w} \cdot \vec{\phi}_{q,d_2} + \vec{w} \cdot \vec{\phi}_{q,d_3} + \vec{w} \cdot \vec{\phi}_{q,d_1}$  which implies  $2(w_1 + w_2) - w_1 > 2(w_1 + w_2)$  and hence  $w_1 < 0$ . But in the previous part of the problem, we showed  $w_1 > 0$  is a requirement for the partial ordering to be satisfied. This is a contradiction and hence there is no such  $\vec{w}$ .
- We could fix this problem by adding additional features to the  $\vec{\phi}_{q,d_i}$  vectors. Another solution is to use a kernel (as a generalization of the orthogonal projection) as in an SVM. This possibility is discussed in [1].
2. (a) “Drunk gets nine months in violin case.” There is a word sense ambiguity in this sentence. The word “case” has two senses. It can refer to a trial or something that holds a violin. The two meanings of this headline can be that a drunk gets nine months in jail after being found guilty in the “violin” trial, or the drunk has to spend nine months in a violin case as a punishment.

- (b) "Stolen painting found by tree."

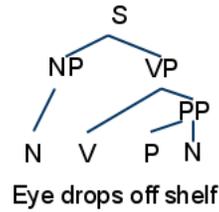
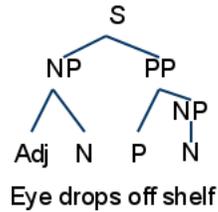
There are two possible solutions to this problem. There can either be a word sense ambiguity or a prepositional phrase attachment ambiguity. In the word sense ambiguity case, the word "by" can have two senses. It can refer to "nearby" or "due to the efforts of."

In the prepositional phrase attachment ambiguity case, the prepositional phrase "by tree" can be attached to either the noun phrase "stolen painting" or the verb "found."

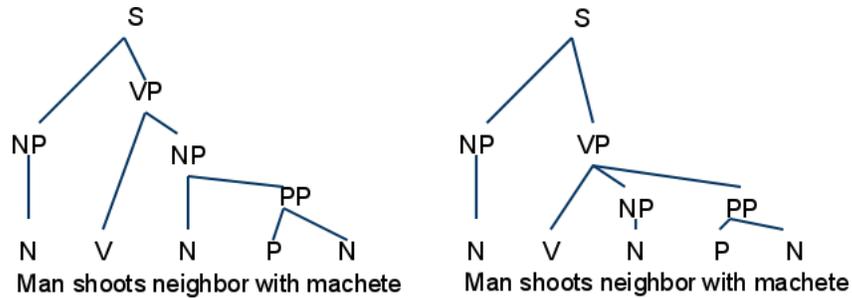
Either way, the two meanings of this headline can be that a person found the stolen painting next to a tree, or the tree found the stolen painting.



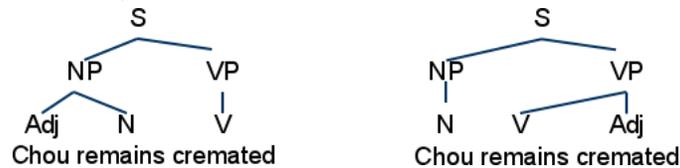
- (c) "Eye drops off shelf." There is part-of-speech ambiguity in this sentence. The word "eye" can either be an adjective that modifies drops or it can be a noun which is in the act of dropping. The word "drops" can either be a noun or a verb. The two meanings of this headline can be that eye drops were taken off the market, or an eye falls off a shelf.



- (d) "Actor sent to jail for not finishing sentence." There is a word sense ambiguity in this sentence. The word "sentence" has two senses. It can refer to a punishment after being found guilty or a sequence of words. The two meanings of this headline can be that an actor gets sent to jail for not completing his punishment, or an actor gets a punishment because he does not finish saying his sentence.
- (e) "Man shoots neighbor with machete." There is prepositional phrase attachment ambiguity in this sentence. The prepositional phrase "with machete" can be attached to either the noun "neighbor" or the verb "shoots." The two meanings of this headline can be that a person shoots his neighbor who was holding a machete, or a person uses a machete to shoot his neighbor.



- (f) “Chou remains cremated” There is part-of-speech ambiguity in this sentence. The word “Chou” can either be an adjective that modifies “remains” or it can be a noun which is still cremated. The word “remains” can either be a noun or a verb, which would make the word “cremated” either a verb or an adjective correspondingly. The two meanings of this headline can be that the remains of Chou were cremated, or that Chou is still cremated.



3. (a) Here are a few ways that we can modify a “standard” English CFG to accommodate headlines. An in-depth analysis of headline grammar can be found in [3].
- Headlines do not generally have determiners. In particular, they almost never have articles (such as “a” or “the”). The noun phrase production for the CFG thus should permit a smaller set of modifiers than that of standard English (and it would omit articles entirely).
  - Headlines lack copulas that indicate class-membership (“to-be” verbs like “is” and “are”). These words would be banished from the CFG.
  - Headlines can be sentence fragments. For example, the headline *Inside the home for angry infants* is a prepositional phrase. Thus the CFG would need a special production like  $S \rightarrow PP$ .
  - Headlines typically omit coordinating conjunctions (e.g. “and,” “but”), but they have other constructs for creating compound sentences that are rare in standard English. For example, there are many headline sentences that begin with a phrase and then a colon, such as *Cleared: The father who killed his drunken neighbour after mistaking him for a burglar*. Thus a grammar would omit coordinating conjunctions and allow these novel colon-based constructs.

- (b) Here are some criteria and techniques from the literature [4].
- i. **Grammatical correctness:** To what extent does the compressed text make sense? One measure is *linguistic score*, the probability of generating a compressed sentence from an n-gram model. This score is used in compression systems to ensure that certain function words are used and to ensure that a sentence's topic words are placed together coherently. Another measure is *SOV Score*, which measures how many of the words in the compressed text are subjects, objects, or verbs. This score is used to ensure that the main parts of a sentence are present in the compression, which is a good indicator of grammatical correctness.
  - ii. **Importance retention:** To what extent do the compressed sentences include the important information from the original text? One measure is the *word significance score*, which is essentially a TF-IDF score. The SOV score is also a measure of importance retention, because subjects, objects, and verbs are often the most significant parts of sentences.
  - iii. **Compression rate:** How short is the compressed text relative to the original? This criterion is especially important for headlines, because there is limited space on the page of a newspaper. One simple measure of compression is the percentage of words removed from the original text.

## 10 References

1. Thorsten Joachims. Optimizing search engines using clickthrough data. KDD, 2002
2. Chiara Bucaria. Lexical and Structural Ambiguity in Humorous Headlines. Youngstown State University. Dept. of English. Master's thesis, 2004.
3. Eva Prášková. Grammar in Newspaper Headlines. University of Pardubice. Department of English and American Studies. Bachelor Paper, 2009.
4. James Clarke, Mirella Lapata, Models for sentence compression: a comparison across domains, training requirements and evaluation measures. Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, p.377-384, July 17-18, 2006, Sydney, Australia Sydney, Australia.