Computationally Identifying Confusing Passages in Textbooks

Amy Zhang

Writing good learning materials is hard.

Oftentimes written by experts who don’t know where novices will get confused.

Hard to quickly tell how confusing a text is likely to be.

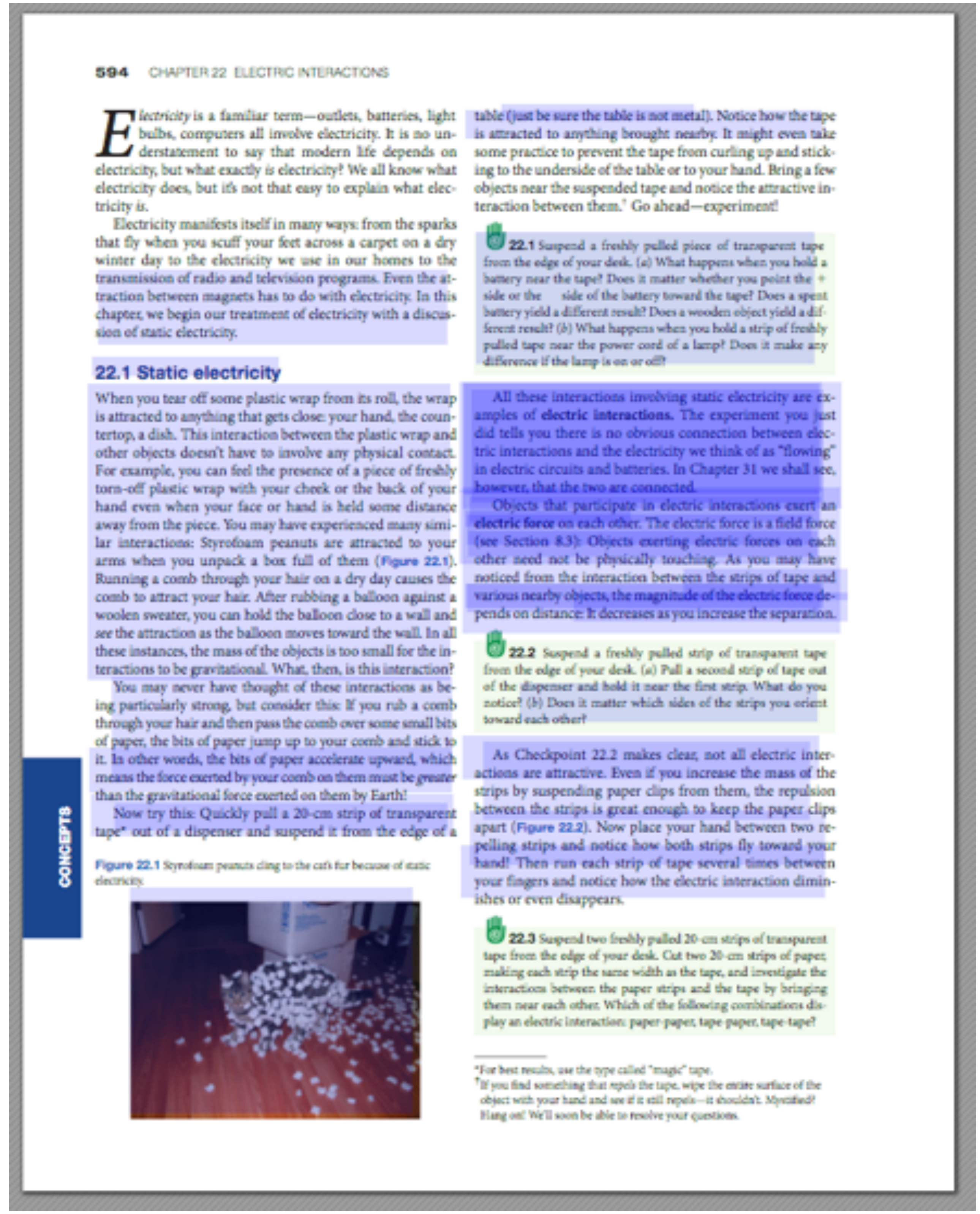
What if we had data that shows where students are confused?

Could we learn characteristics about what makes text confusing and use it to predict confusing sections of textbooks?

- We use data from **NB** - a PDF annotation software used by 100's of classes
- NB lets students highlight parts of text and leave comments for discussion
- Here we focus on comments that express **confusion** (determined using heuristics such as presence of "?")

Example NB comment:

"So in figure 22.3 a top layer tape is pulled from a bottom layer tape that is flat on a table. In 22.5 the top layer is pulled off from the bottom layer while hanging on an edge. What is it exactly that the table does that gives it a different outcome than if the two pieces of tape were hanging? This parts just confusing me"



594 CHAPTER 22 ELECTRIC INTERACTIONS

Electricity is a familiar term—outlets, batteries, light bulbs, computers all involve electricity. It is no understatement to say that modern life depends on electricity, but what exactly *is* electricity? We all know what electricity does, but it's not that easy to explain what electricity is.

Electricity manifests itself in many ways: from the sparks that fly when you scuff your feet across a carpet on a dry winter day to the electricity we use in our homes to the transmission of radio and television programs. Even the attraction between magnets has to do with electricity. In this chapter, we begin our treatment of electricity with a discussion of static electricity.

22.1 Static electricity

When you tear off some plastic wrap from its roll, the wrap is attracted to anything that gets close: your hand, the countertop, a dish. This interaction between the plastic wrap and other objects doesn't have to involve any physical contact. For example, you can feel the presence of a piece of freshly torn-off plastic wrap with your cheek or the back of your hand even when your face or hand is held some distance away from the piece. You may have experienced many similar interactions: Styrofoam peanuts are attracted to your arms when you unpack a box full of them (Figure 22.1). Running a comb through your hair on a dry day causes the comb to attract your hair. After rubbing a balloon against a woolen sweater, you can hold the balloon close to a wall and see the attraction as the balloon moves toward the wall. In all these instances, the mass of the objects is too small for the interactions to be gravitational. What, then, is this interaction?

You may never have thought of these interactions as being particularly strong, but consider this: If you rub a comb through your hair and then pass the comb over some small bits of paper, the bits of paper jump up to your comb and stick to it. In other words, the bits of paper accelerate upward, which means the force exerted by your comb on them must be greater than the gravitational force exerted on them by Earth!

Now try this: Quickly pull a 20-cm strip of transparent tape* out of a dispenser and suspend it from the edge of a table (just be sure the table is not metal). Notice how the tape is attracted to anything brought nearby. It might even take some practice to prevent the tape from curling up and sticking to the underside of the table or to your hand. Bring a few objects near the suspended tape and notice the attractive interaction between them.† Go ahead—experiment!

22.1 Suspend a freshly pulled piece of transparent tape from the edge of your desk. (*a*) What happens when you hold a battery near the tape? Does it matter whether you point the + side or the − side of the battery toward the tape? Does a spent battery yield a different result? Does a wooden object yield a different result? (*b*) What happens when you hold a strip of freshly pulled tape near the power cord of a lamp? Does it make any difference if the lamp is on or off?

All these interactions involving static electricity are examples of **electric interactions.** The experiment you just did tells you there is no obvious connection between electric interactions and the electricity we think of as "flowing" in electric circuits and batteries. In Chapter 31 we shall see, however, that the two are connected.

Objects that participate in electric interactions exert an **electric force** on each other. The electric force is a field force (see Section 8.3): Objects exerting electric forces on each other need not be physically touching. As you may have noticed from the interaction between the strips of tape and various nearby objects, the magnitude of the electric force depends on distance: It decreases as you increase the separation.

22.2 Suspend a freshly pulled strip of transparent tape from the edge of your desk. (*a*) Pull a second strip of tape out of the dispenser and hold it near the first strip. What do you notice? (*b*) Does it matter which sides of the strips you orient toward each other?

As Checkpoint 22.2 makes clear, not all electric interactions are attractive. Even if you increase the mass of the strips by suspending paper clips from them, the repulsion between the strips is great enough to keep the paper clips apart (Figure 22.2). Now place your hand between two repelling strips and notice how both strips fly toward your hand! Then run each strip of tape several times between your fingers and notice how the electric interaction diminishes or even disappears.

22.3 Suspend two freshly pulled 20-cm strips of transparent tape from the edge of your desk. Cut two 20-cm strips of paper, making each strip the same width as the tape, and investigate the interactions between the paper strips and the tape by bringing them near each other. Which of the following combinations display an electric interaction: paper-paper, tape-paper, tape-tape?

Figure 22.1 Styrofoam peanuts cling to the cat's fur because of static electricity.

*For best results, use the type called "magic" tape.

†If you find something that repels the tape, wipe the entire surface of the object with your hand and see if it still repels—it shouldn't. Mystified? Hang on! We'll soon be able to resolve your questions.

CONCEPTS

Data Cleaning

- We pick 10 chapters from an introductory Physics textbook used in several classes.
- Each chapter has been annotated by 3 or 4 sections
- We split text into **paragraphs** and annotate them as confusing if 3/4 sections had a confusing comment.
- In total: 929 paragraphs, 596 not confusing

Example Paragraphs from the Same Page

Confusing

A discharged tape strip interacts in the same way as objects that carry no charge. Such objects are said to be electrically neutral. They do not interact electrically with other neutral objects, but they do interact electrically with charged objects. We shall examine this surprising fact in more detail in Section 22.4.

Not Confusing

Where does the electrical charge on a charged tape strip come from? Is charge created when two strips are pulled apart as in Figure 22.3? This is something we can check by sticking two strips of tape together, rubbing with our fingers to remove all charge from the combination, and then quickly separating the two strips (Figure 22.5).

We also have data about the book and each page: figures and tables on the page, bolded text (summaries, definitions), position of paragraph.

Features

- Unigrams, bigrams
- Type of paragraph: num summaries, definitions, part of a list
- Page structure: page num, num figures on the page
- Length: num sentences, avg length of sentence
- Technical: num equations, num variables, num numbers, num values
- Vocabulary: first occurrence of vocab word, max num pages since last occurrence of vocab word
- LIWC: affect, discrepancy, cognitive processes, tentative
- Part of speech: adjectives, numerals, nouns
- Non-vocab words: first occurrence of non-common non-vocab words

Results (avg over 10-fold cross validation)

	Acc	Prec	Recall	F1
Random (stratified):	.54	.36	.36	.36
All Not Confusing:	.64	0	0	0
Unigram (SVM):	.49	.30	.29	.29
Unigram tf-idf (SVM):	.47	.26	.22	.22
Unigram (MaxEnt):	.50	.31	.27	.27
Unigram tf-idf (MaxEnt)	.53	.14	.07	.08
Unigram Naive Bayes:	.49	.31	.33	.28
Unigram NB tf-idf:	.61	.05	.03	.04
Ling Model (SVM):	.70	.64	.42	.49
Ling Model (RF):	.70	.66	.41	.49
Ling Model (MaxEnt):	**.73**	**.71**	**.51**	**.57**

Results by Feature Group

(Best MaxEnt model, avg over 10-fold cross validation)

	Acc	Prec	Recall	F1
Paragraph content:	.70	.71	.30	.42
Page Structure:	.69	.62	.50	.53
LIWC:	.65	.71	.05	.09
Vocab words:	.66	.68	.13	.21
Text length:	.64	.51	.06	.11

Discussion

- Bag-of-words models don’t perform well, perhaps because they take too much of the topical content into account

What kinds of texts are confusing?

- Shorter paragraphs but longer sentences and longer words
- Less numbers, variables, and equations, but more adjectives and use of words with affect, tentativeness, discrepancy
- Earlier paragraphs in the chapter, less figures on the page
- Introducing new vocab, summarizations, more pages between last occurrence of vocab word
- Introducing new uncommon terms without signposting them
- Using more common but potentially imprecise domain-specific terminology

Future Work

- Predict pairs of confusing/not-confusing passages on the same page or within a page of each other.

- Non-binary methods such as regression or multi-class classification to get at degree or ranking of confusion.

- Classify passages into types – example, explanation, claim, introduction, summary. How does their ratio or sequence affect confusion? What differentiates confusion within the types?

- Can we use the thread discussion data to then provide suggestions to authors?