ANALYSIS ON PUNCTUATIONS IN ONLINE REVIEWS

TONGZHOU WANG   CS C281A FINAL PROJECT

OVERVIEW

- Explore information in punctuations in Steam online reviews.
- Two different directed graphical models are analyzed using a publicly available dataset.
- Gather insights about punctuations and structures in online reviews.

MOTIVATION

Many text analyses focus mainly on the words. However, punctuations in texts obviously also carry huge amount of information. In fact, using punctuations as features has advantages including:

1. insensitive to typos,
2. less sensitive to language used,
3. expressing strong emotions, and
4. limited in types, resulting in much smaller sized models.

Want to investigate the relations among punctuations in Steam online game review dataset.
Aforementioned advantages of punctuation features are particularly useful in this dataset because online game reviews

1. tend to have many typos,
2. can be in many different languages,
3. sometimes express strong emotions, e.g. anger, excitement, etc.

METHODS

- Preprocess data s.t. each sentence ends in proper punctuation.
- Separate the entire review dataset into positive ones and negative ones.
- Fit two models on positive data and negative data respectively.
- Analyze different punctuation relations in positive and negative data.

MODEL

Sentences in reviews usually serve for different purposes, e.g. introduction, listing pros & cons, scoring, etc.

Directed model with sentence type:

1. Each sentence are considered as having one of m sentence types.
2. Sentences types in a review are from a Markov chain (π_s, A_s).
3. For each sentences, given its sentence type s_i, punctuations are from Markov chain $(\pi_p^{s_i}, A_p^{s_i})$.

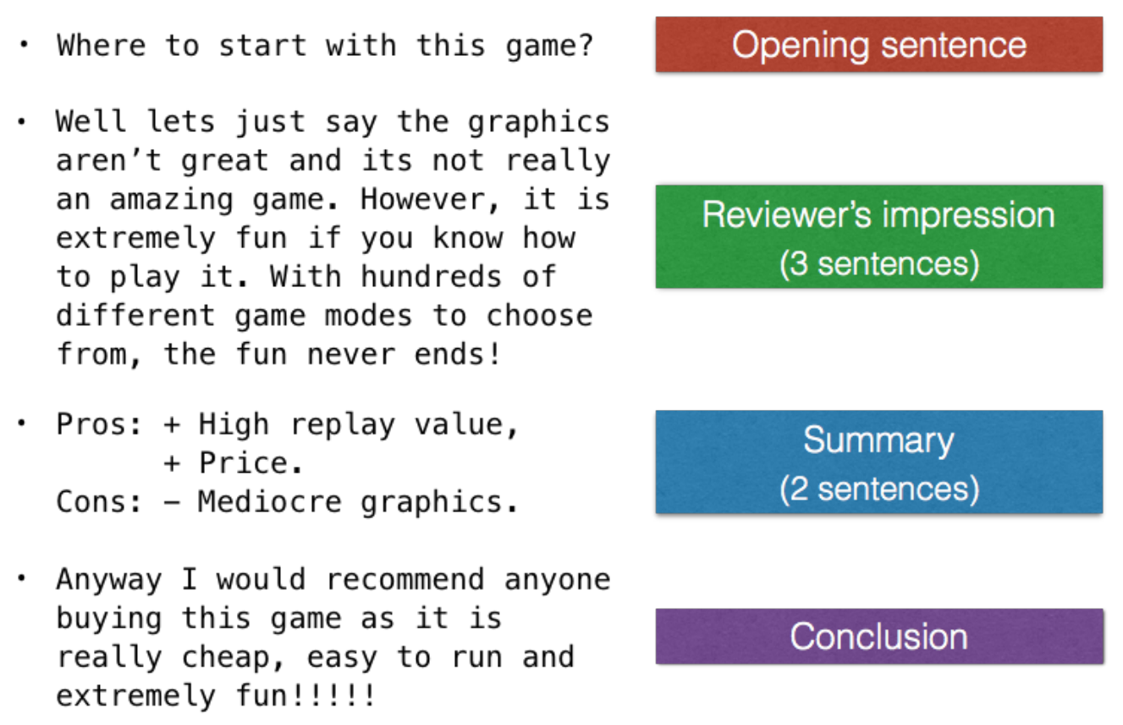
This model

- Essentially an HMM but with P(punctuations|sentence type) parametrized by a Markov chain
- Can use EM algorithm to approximately find MLEs.
- Can use Forward-backward algorithm for hidden variable marginals in EM.

FUTURE WORK

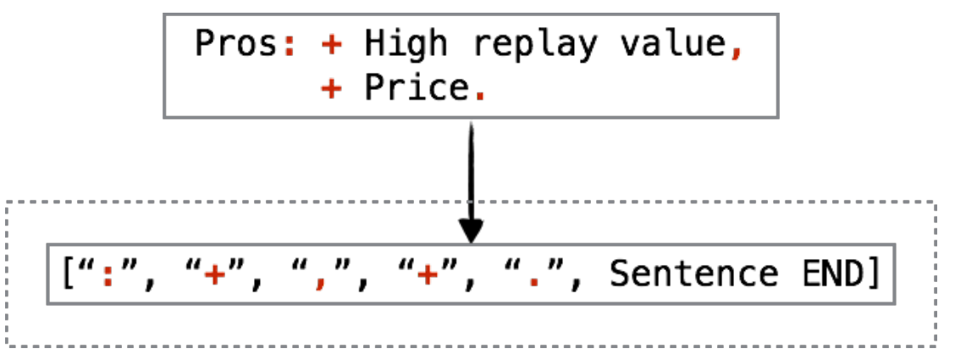
1. Positive/negative prediction using the two fitted models.
2. Experiment and analyze best choice for m number of sentence types.

VISUALIZATIONS AND RESULTS



- Where to start with this game?
- Well lets just say the graphics aren't great and its not really an amazing game. However, it is extremely fun if you know how to play it. With hundreds of different game modes to choose from, the fun never ends!
- Pros: + High replay value,
  + Price.
  Cons: – Mediocre graphics.
- Anyway I would recommend anyone buying this game as it is really cheap, easy to run and extremely fun!!!!!

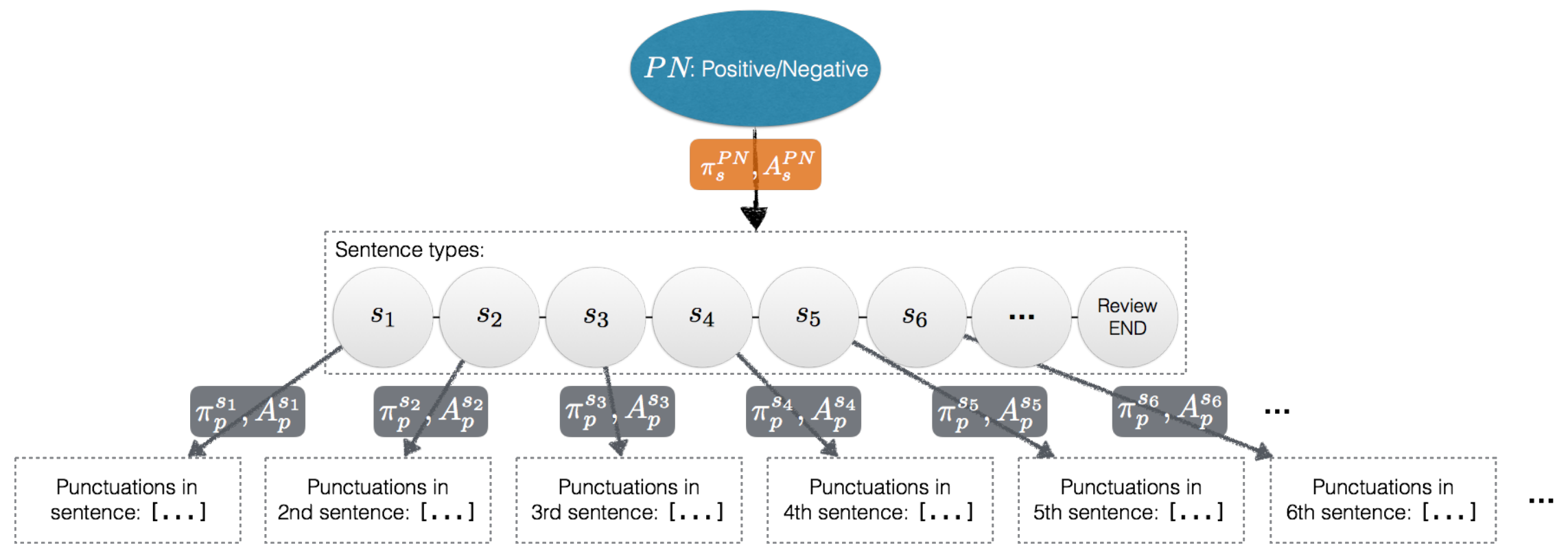
(a) Hypothetical sentence types in a review example



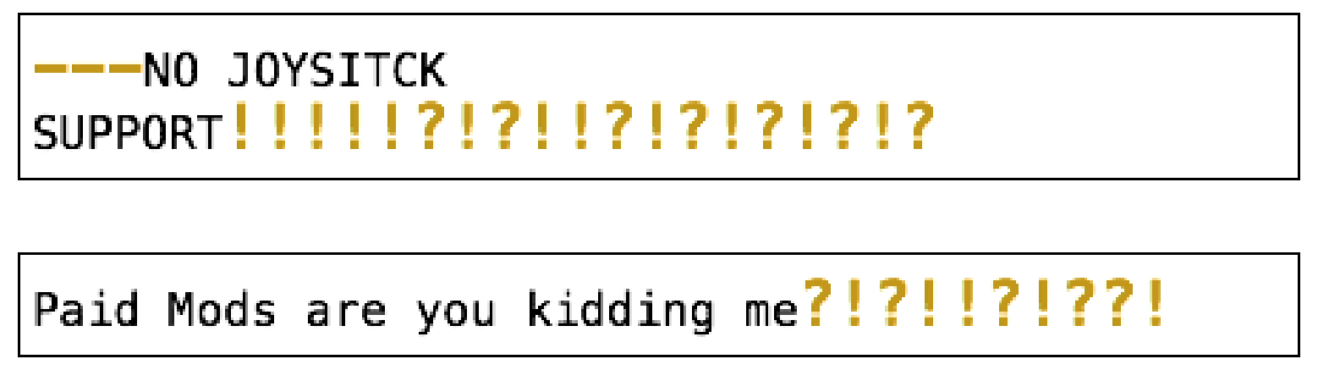
```
Pros: + High replay value,
      + Price.
```

```
[“:”, “+”, “,”, “+”, “.”, Sentence END]
```

(b) Punctuation data example of a sentence



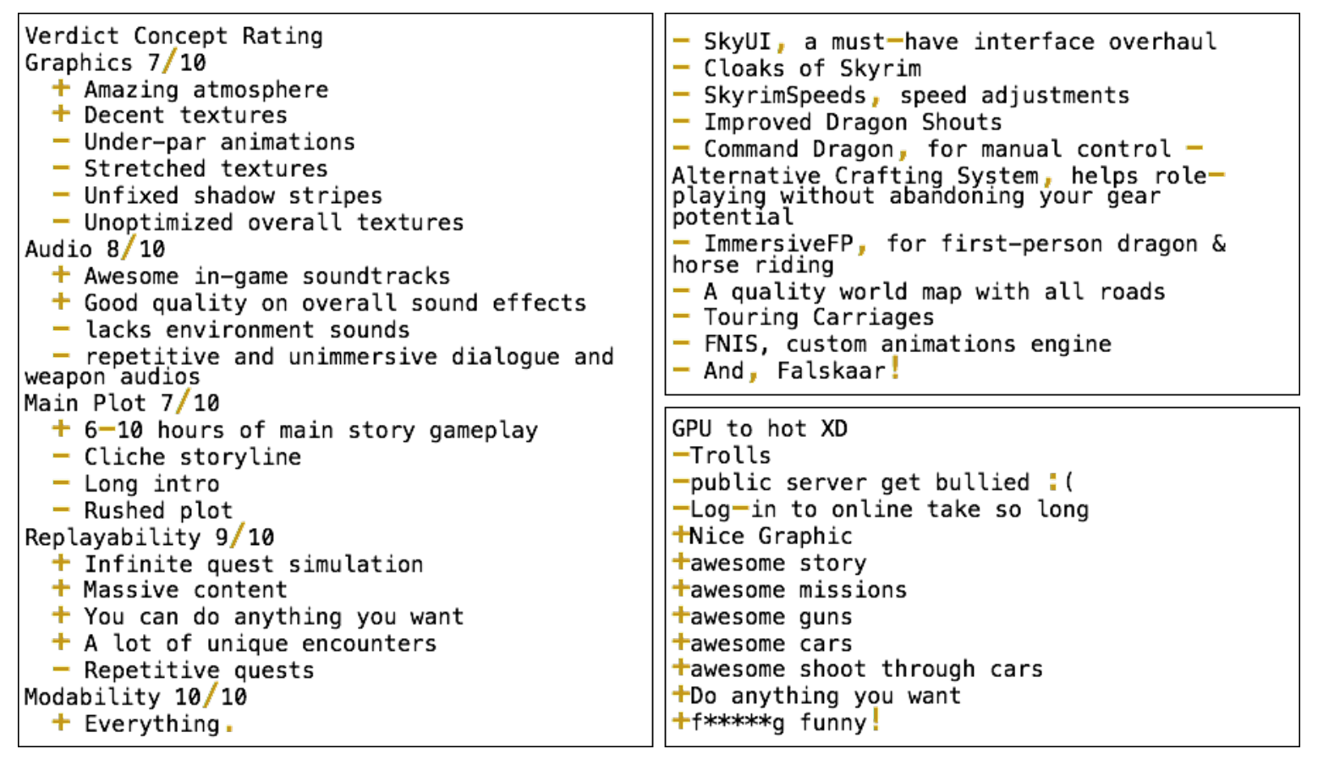
(c) Model visualization



```
---NO JOYSITCK
SUPPORT!!!!!?!?!!?!?!?!?!?
```

```
Paid Mods are you kidding me?!?!!?!??!
```

(d) Sentence type **1** of **negative** reviews examples

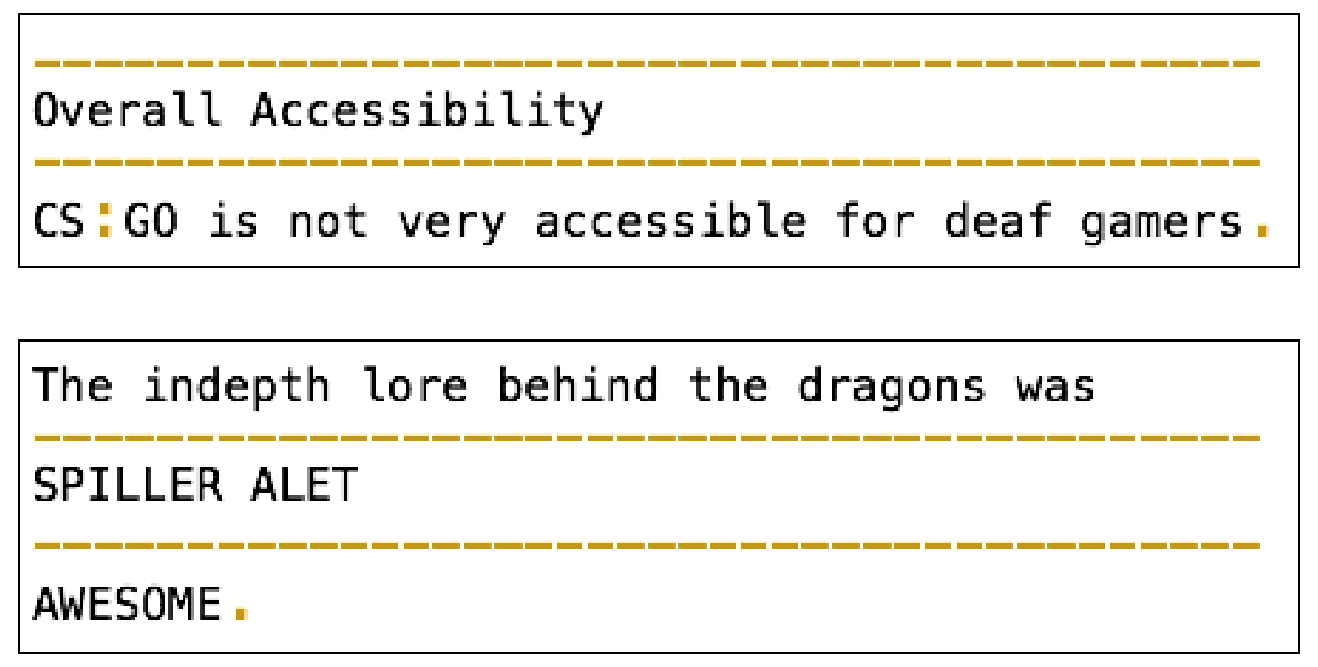


```
Verdict Concept Rating
Graphics 7/10
  + Amazing atmosphere
  + Decent textures
  - Under-par animations
  - Stretched textures
  - Unfixed shadow stripes
  - Unoptimized overall textures
Audio 8/10
  + Awesome in-game soundtracks
  + Good quality on overall sound effects
  - lacks environment sounds
  - repetitive and unimmersive dialogue and
weapon audios
Main Plot 7/10
  + 6-10 hours of main story gameplay
  - Cliche storyline
  - Long intro
  - Rushed plot
Replayability 9/10
  + Infinite quest simulation
  + Massive content
  + You can do anything you want
  + A lot of unique encounters
  - Repetitive quests
Modability 10/10
  + Everything.
```

```
- SkyUI, a must-have interface overhaul
- Cloaks of Skyrim
- SkyrimSpeeds, speed adjustments
- Improved Dragon Shouts
- Command Dragon, for manual control -
Alternative Crafting System, helps role-
playing without abandoning your gear
potential
- ImmersiveFP, for first-person dragon &
horse riding
- A quality world map with all roads
- Touring Carriages
- FNIS, custom animations engine
- And, Falskaar!
```

```
GPU to hot XD
-Trolls
-public server get bullied :(
-Log-in to online take so long
+Nice Graphic
+awesome story
+awesome missions
+awesome guns
+awesome cars
+awesome shoot through cars
+Do anything you want
+f*****g funny!
```

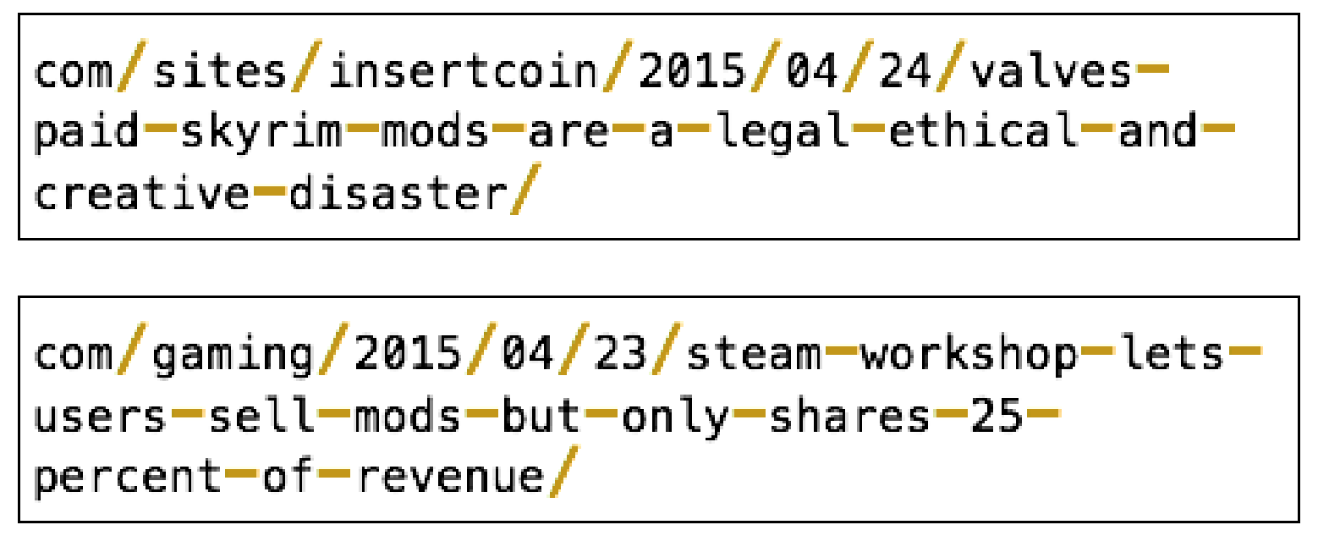
(e) Sentence type **2** of **positive** reviews examples



```
-----------------------------------------
Overall Accessibility
-----------------------------------------
CS:GO is not very accessible for deaf gamers.
```

```
The indepth lore behind the dragons was
-----------------------------------------
SPILLER ALET
-----------------------------------------
AWESOME.
```

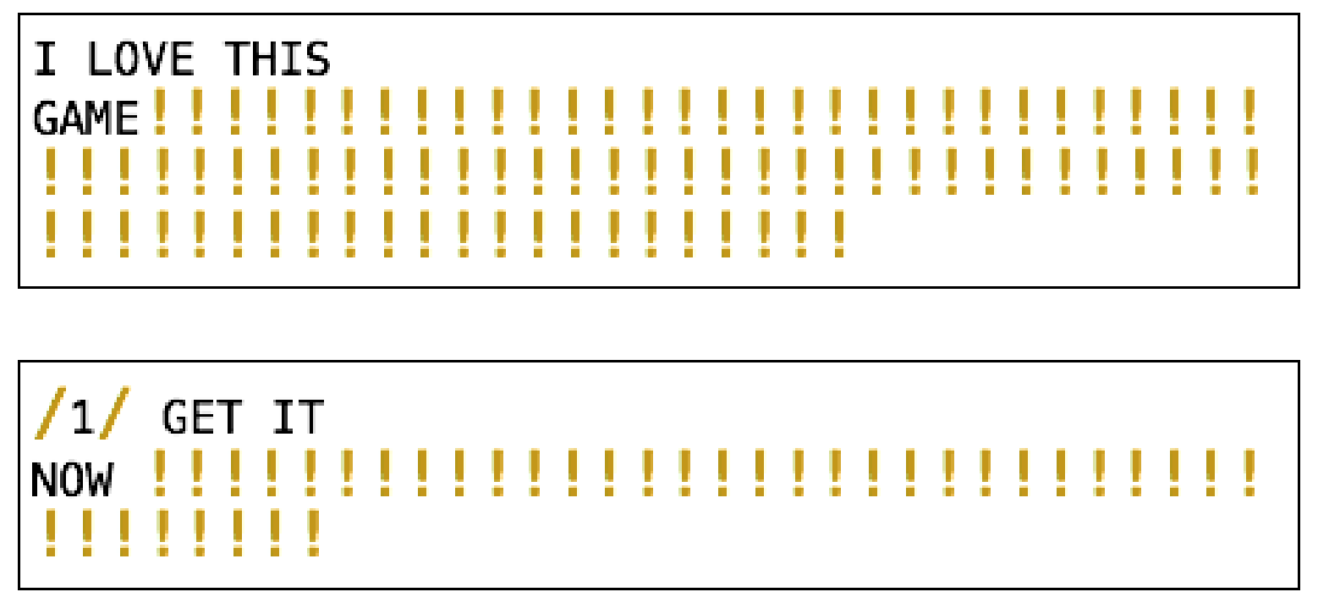
(f) Sentence type **4** of **positive** reviews examples



```
com/sites/insertcoin/2015/04/24/valves-
paid-skyrim-mods-are-a-legal-ethical-and-
creative-disaster/
```

```
com/gaming/2015/04/23/steam-workshop-lets-
users-sell-mods-but-only-shares-25-
percent-of-revenue/
```

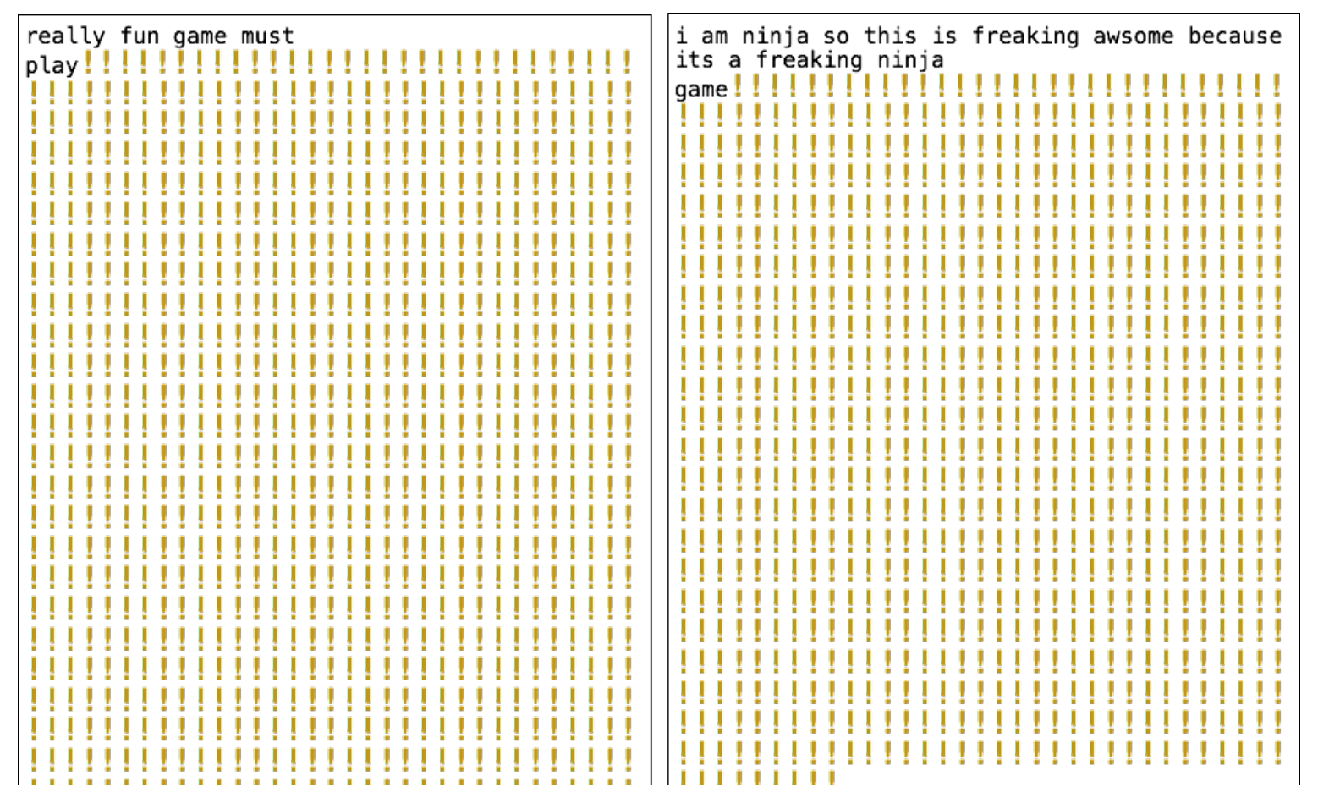
(g) Sentence type **0** of **negative** reviews examples



```
I LOVE THIS
GAME!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!!!!
```

```
/1/ GET IT
NOW !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!
```

(i) panel text:



```
really fun game must
play!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
```

```
i am ninja so this is freaking awsome because
its a freaking ninja
game!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
```

(h) Sentence type **1** of **positive** reviews examples

(i) Sentence type **5** of **positive** reviews examples

OBSERVATIONS AND ANALYSIS

- Punctuation data are very informative.
- Sentence types successfully capture punctuation patterns:
  - **Negative** type **1**: pattern of anger.
  - **Positive** type **2**: pros and cons.
  - **Positive** type **4**: formatting pattern.
- Transition probability also gives interesting insights:
  - **Negative** reviews have high prob. from type **0** (url pattern) to REVIEW END.
  - **Positive** reviews have high prob. from type **1** (!!!) to type **5** (more !!!!!!).

RESOURCES AND REFERENCES

[1] Mulholland, *Steam Review Datasets*, https://github.com/mulhod/steam_reviews.

[2] Gu, Leon, *EM and HMM*.