Testing Interactive Narratives using Personality Types

David Pereira
Instituto Superior Técnico
Lisbon, Portugal
davidgarcia@tecnico.ulisboa.pt

ABSTRACT

Research has shown people are more receptive to information presented from the perspective of their own personality features, and that it is possible to predict a person's Myers-Briggs personality type from their writing style.

I developed a tool that simulates what options people of different personality types select when playing through text-based interactive narrative games. The tool uses text classifiers to analyze the options based on the writing style of their text and selects the most adequate ones.

The user tests were not conclusive as the volunteers didn't pick the options I had anticipated. Amongst other possibilities, this could mean that a person's Myers-Briggs personality type may not be a good indicator of what options they're more likely to select, or that people may not be that influenced by the writing style of the selectable options.

Keywords - Interactive Narratives, Playtesting, Text Mining, Myers-Briggs Type Indicator

1. Introduction

Research [1] has shown that some players report an increase in their level of enjoyment of computer role-playing games when the story is adapted to their learned preferences. That increase was noticed especially on their level of entertainment and agency when comparing adaptive stories to fixed ones.

One possible way of catering games to individuals is by focusing on the player's playing style and correlating it with pre-defined profiles to customize their experience. That way, along with reaping the benefits of having a larger audience, game writers could predict how different types of players would play their game, simplifying the process of game creation, and of playtesting by reducing the dependency on human players.

One thing that influences a player's actions is their personality. Research has shown that it is possible to learn one's personality type from their writing style [6], and that people are more sympathetic towards information presented from the perspective of their own personality features (i.e. introverts favor messages written from an introvert's perspective) [5].

In this thesis, I propose that, when analyzing an interactive narrative game, it is possible to predict what options a player is more likely to select based on their personality type. I created a tool that simulates those selections by extracting the options' personality types using text classifiers.

The tool's main objective is to provide game writers with gameplays of their story customized to personality types of their choice, to potentially help authors expand their stories to make them more enjoyable to people of different personality types. The tool's main features are the following:

- Receive an interactive narrative game and a Myers-Briggs personality type
- Provide a short report containing the best paths for the selected personality type, and the affinity scores between each step of the paths and the selected personality type, and the overall score of the path

The goal of the tool is not to improve the story's theoretical enjoyment directly; it is a mere guide authors could use to learn the affinity scores of their stories to specific personality types – whether they choose to act on that is entirely up to them and outside of the capabilities of the tool.

To test the tool, I used it to simulate gameplays of an interactive narrative game for each personality type and compared the results to gameplays of that same game by people of those personality types.

The tests were not conclusive and seemed to suggest there's no significance to the options picked by the volunteers. This could be an indication that a person's Myers-Briggs personality type may not be a good way of predicting what options they're more likely to select, or that people may not be influenced by writing style of the selectable options.

2. Related Work

2.1. Interactive Narrative Games

Interactive storytelling [1] involves narratives in which the sequence of events experienced by the player is based on their interactions with the story world, allowing the player to reach different outcomes.

Twine [7] is an interactive story generator released in 2009 by web developer and game designer Chris Klimas.

Twine stories are divided into Passages, which contain a title, a tag (optional), and a body. The body of a Passage can contain text, blocks of code, and zero or more Links. A Link is selectable option that takes the player to another Passage. Authors create Links by typing double brackets ("[[" and "]]") around text.

2.2. Player Modelling

Player modeling [1] [9] [25] is the creation of computational models of players based on their tendencies and behavioral patterns, with the intent of predicting how they would behave in certain situations and under certain conditions.

PaSSAGE [1] is an interactive storytelling system that learns the playing style of the player and adapts the story of the game accordingly. The system borrows the player types suggested by Robin D. Law in *Robin's Laws of Good Game Mastering* [10]:

- Fighters (F)
- Method Actors (M)
- Storytellers (S)
- Tacticians (T)
- Power Gamers (P)

PaSSAGE uses a system of weights to learn the player's model. The higher the weight, the stronger the model's belief that that is the preferred playing style. The system of weights is organized in a vector, such as the following at the start:

(**F**=1, **M**=1, **S**=1, **T**=1, **P**=1)

As the player performs actions, the weights system updates, with different actions having different impacts. For example, if the player performs an action of a Method Actor, M's value in the vector increases:

(**F**=1, **M**=10, **S**=1, **T**=1, **P**=1)

The second study [11] focused on constructing models of players for Tomb Raider: Underworld, based on data obtained during gameplay (completion time, number of times the player asked for a hint or answer to a puzzle, total number of deaths, further divided by cause of death).

After processing the evaluated data using Emergent Self-Organizing Maps (ESOM) [12], the researchers obtained four clusters of playing behavior.

The main difference between these two approaches is that while the first attempts to fit the player into an already existing model, the second creates player models from scratch. The player models of the second option are expected to be more accurate since they reflect how people *actually* played the game, while the weights of the first approach may not be well calibrated.

2.3. Automated Playtesting

While developing a game, designers need to be aware of the scope of possible actions and outcomes. As the development progresses, the game also grows in complexity, becoming harder to control the scope of all possible scenarios that result from different interactions. For this reason, playtesting can be used to search for cases the designer didn't account for, such as exploits and fail states.

Playtesting is also important in terms of providing players with a balanced experience. Game balance is particularly important in terms of making sure all players start with equal chances of winning, to attract as many players as possible, but also in terms of rewarding plays for improving their skills, since games who manage to do that have more success in retaining players [13].

However, apart from requiring human testers, playtesting also requires collecting, treating and interpretating huge quantities of data, meaning it is a very expensive process – one that not all game developing companies, especially indie ones, can afford [14].

In one study [15], the researchers used artificially intelligent personas to test the game MiniDungeons 2.

The researchers used a variation of the MCTS algorithm to simulate the behavior of the different archetypes.

In terms of results, the personas played the game more efficiently than the baseline agents, while still maintaining the differencing in-game metrics that made them unique.

This method is a more budget-friendly alternative to simulate the behavior of human testers. It can also be used whenever human feedback isn't readily available, or when it might not make financial sense to test with humans, such as when testing small changes.

In another research [16], the authors used deep learning to predict the next move in a Candy Crush Saga playout.

The training of the network was done with player data from previous levels, with the tool learning the most common actions taken by players in similar states. Correlation with average level difficulty was increased, resulting in more accurate move predictions, and requiring less computation time.

Unlike the first approach, this one requires player data. Consequently, it is a more expensive alternative, though it is more reliable. Still, it can potentially be used to playtest at a later stage of development, when there's not enough time to test with human players, such as in the case of a last-minute addition, or when the designers only need to test small changes.

2.4. Text Mining for Personal Characteristics

Text mining [17] is a process of analyzing natural language text with the intent of detecting lexical or linguistic patterns to extract information.

A possible application of text mining is sentiment analysis. In the context of text mining, sentiment analysis is the process of detecting positive or negative sentiment in text.

For one study [20], the authors created a tool to classify tweets into positive or negative sentiments. The input consisted of tweets labeled by humans as "positive," "negative," or "neutral".

The tool reached an accuracy of 75.39% when the following features were taken into consideration:

- Number of negation, positive, and negative words
- Number of extremely positive, extremely negative, positive, and negative emoticons
- Number of positive and negative hashtags, capitalized words, and exclamation words
- Sum of the polarity scores of all the words

In another research [21], the authors focused on the dichotomy introversion / extraversion.

From a set of tweets and an indication of whether they were "extroverted" or "introverted," the authors created a tool that could tell whether a text was "extroverted" or "introverted" with an accuracy of 84.07%.

2.5. Myers-Briggs Type Indicator

The Myers-Briggs Type Indicator (MBTI) [22] is a personality classification system that attempts to classify people into 4-letter acronyms, based on how they perceive the world and make decisions.

Each letter of the acronym comes from the preferred quality of each of the following dichotomies:

Dichotomy 1	**E**xtraversion	**I**ntroversion
Dichotomy 2	**S**ensing	i**N**tuition
Dichotomy 3	**T**hinking	**F**eeling
Dichotomy 4	**J**udging	**P**erceiving

Table 1: Myers-Briggs dichotomies

For example, someone whose preferred qualities are **I**ntroversion, i**N**tuition, **T**hinking, and **J**udging is an **INTJ**.

In one study [23], the authors aimed to predict a person's MBTI classification based on their social media posts.

The researchers created a binary classifier for each of the dichotomies (E/I, S/N, T/F, J/P), with the aggregated the results forming the four-letter acronym. The results obtained were the following:

Function	E / I	S / N	T / F	J / P
Prediction Accuracy	89%	89%	69%	67%

Table 2: Accuracies achieved for each Myers-Briggs dichotomy by Deep Learning.

Another research [24] aimed to construct semantic representations and use them to identify the different personality types. The input consisted of tweets and their creator's personality type. They obtained the following results:

Function	E / I	S / N	T / F	J / P
Prediction Accuracy	84.9%	88.4%	87.0%	78.8%

Table 3: Accuracies achieved for each Myers-Briggs dichotomy by the SVM (Support Vector Machine).

3. The Tool

3.1. Origins

My idea was to mix the concept of those memes with interactive storytelling: I wanted to simulate how the different personality types would play through an interactive narrative game.

This could help authors plan their stories to better accommodate the preferences of the different Myers-Briggs personality types, aiding in the creation of a more personalized game experience.

The usefulness of this tool, of course, rested on the premise that people made different choices based on their personality type, that different people of the same personality type made similar choices, and that people were more likely to select options that were presented from the perspective of their own personality features [5], which I assumed was also valid for options presented from the perspective of their Myers-Briggs dichotomies.

3.2. Basic Concept

From an interactive narrative game and one or more Myers-Briggs personality types, my tool produces a short report containing the best gameplay paths for each of the selected personality types.

To determine the best gameplay paths, at each decision point the tool analyzes the selectable options with the help of external text classifiers and selects the best-fitting one. Essentially, the tool uses Myers-Briggs personality types as player archetypes.

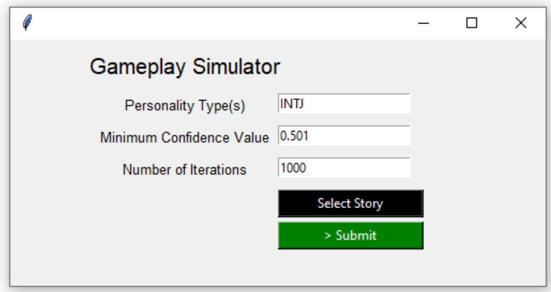


Figure 1: The tool's interface with an example input

Input

3.2.1. Stories

The interactive narrative game must be in Entweedle 1.0.3 format, which can be obtained through Twine.

There can't be any cycles, custom scripts or variables in the story, and the starting passage must be identified with the tag 'START'. The accepted formatting of a passage is as follows:

::Passage Title [Tag] {"position":"xx,yy","size":"aa,bb"}
Message
[[Option A | Link A]]
[[Option B | Link B]]

"Tag" and "Message" are optional, and the content of "position" and "size" is irrelevant. "Option A" and "Option B" are the options that players can see, and "Link A" and "Link B" the destinations those options lead to. There is no limit to the number of options.

3.2.2. Personality Types

The personality types are those of the MBTI. The tool can receive multiple personality types at once, separated by a space.

The personalities can make use of 1 to 4 dichotomies, allowing the analysis of the different dichotomies separately, as the tool reorganizes the letters and replaces any invalid or missing ones with "x." If the user enters two letters corresponding to the same dichotomy, only the first is considered.

3.2.3. Minimum Confidence Value

The user also has the option to enter a minimum value for their MBTI letters to be considered valid. This will be explained in more detail in section 3.4.2.

3.2.4. Number of Iterations

How many times to run the algorithm. Running it more times increases the chances of showing all possible best paths.

3.3. Tool Functioning

3.3.1. Importing the Story

The tool analyzes the input story and divides passages and links into objects of two classes: Passage and Link, respectively.

A Passage object stores the title of the passage, its tag, and a list of Links. A Link object stores the title of the link, its personality type, parent passage, and destination passage. Its personality type is calculated when the story is imported, using the method described in section 3.4.2. Any other information from the original story is discarded.

3.3.2. Personality Extraction

To classify a portion of text, the tool uses four publicly available text classifiers which evaluate, respectively, the four dichotomies of the MBTI, based on the text's writing style:

- **E**xtraversion *versus* **I**ntroversion [19]
- **S**ensing *versus* i**N**tuition [18]
- **T**hinking *versus* **F**eeling [8]
- **J**udging *versus* **P**erceiving [4]

Each classifier returns the degree of certainty associated with each element of the dichotomy. From there, the tool generates the input text's personality type.

For the input "What do we do now?", the tool computes the personality type ISFP.

Dichotomy	Strongest
Extraversion (E) / Introversion (I)	I – 98%
Sensing (S) / Intuition (N)	S – 77%
Thinking (T) / Feeling (F)	F – 65%
Judging (J) / Perceiving (P)	P – 74%

Table **4**: Information provided by the classifiers for the input "What do we do now?"

The user can enter a minimum value (Minimum Confidence Value) for an option's certainty to be considered valid. Taking the previous example, with a minimum confidence value of 0.7, "What do we do now?" would be interpreted as ISxP instead of ISFP.

If the user sets the Minimum Confidence Value to 0, in case of a tie (for example, Strongest = P – 50%), the tool selects one of the elements of the dichotomy at random.

3.3.3. Score Calculation

The score of a Link is the proportion of matching letters between the input personality type and the link's personality type obtained via 3.4.2, ignoring "x"s of the input personality type.

Input Personality	Link Personality	Score
ISFP	**I**xTJ	0.250
E**NFP**	I**NFP**	0.750
I**N**xP	E**N**TJ	0.333

Table 5: Examples of matching scores.

3.3.4. Selection Algorithm

The tool, starting from the passage with the tag "START," plays the game from passage to passage by selecting a link to transport it to the next passage, until the end of the game. The tool has 2 different modes of choosing what links to select:

3.3.4.1. Best Immediate Score

The tool chooses the link with the highest score without any consideration for future passages or links. In case of a tie, it selects one of the best options at random.

3.3.4.2. Best-First Search

The tool uses a best-first search to select the best path from the first passage to all the end passages and picks the path with the highest score.

3.3.4.3. Discussion

I considered these two steps necessary because the first simulates how humans make their selections - unaware of future choices - while the second method finds the *absolute* best path.

The second method covers both linear stories – where the links selected by the player don't affect the next available options – as well as branching narratives, where future available options are based on previous options.

The Best-First Search mode was developed only to provide authors with more information about their stories; the Best Immediate Score mode is the focus of this thesis.

3.3.5. Total Score

It's the average score of the selected links' individual scores, excluding links that were the only selectable option.

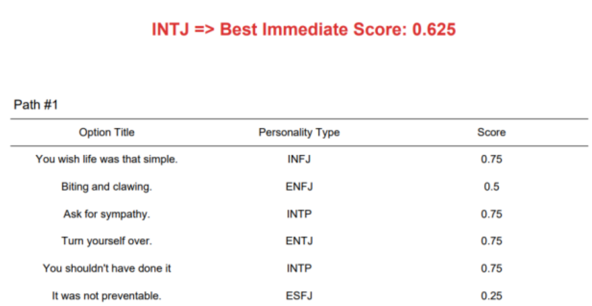
3.3.6. Tool in Action

For each personality type entered, for each passage in the story, the tool analyses all possible options and selects one, until an ending – a decision point without options to choose from – is reached. In the end, the tool returns a short report with all the chosen options.

3.3.7. Results

For each selected personality type, the analysis consists of the maximum score achievable by that personality type, followed by the options selected to achieve that score. If there are multiple paths to achieve that same maximum, the analysis contains all of them.

If the best-first score is higher than the best immediate score, the analysis contains both scores and respective best paths, otherwise it just contains the best immediate score and its path(s).



INTJ => Best Immediate Score: 0.625

Path #1

Option Title	Personality Type	Score
You wish life was that simple.	INFJ	0.75
Biting and clawing.	ENFJ	0.5
Ask for sympathy.	INTP	0.75
Turn yourself over.	ENTJ	0.75
You shouldn't have done it	INTP	0.75
It was not preventable.	ESFJ	0.25

Figure **2**: Example of a short report

3.3.8. Limitations of the tool

The main issue with the tool is its dependence on the personality classification tools.

Another negative aspect is the fact the first run of a story takes a long time to process due to the personality classification tools. However, the tool keeps the association between each option and its personality type between runs.

Another negative aspect comes down to the import format limitations. They limit the content of the stories, for example stories where the player can undo certain actions or perform the same action multiple times (cycles) aren't possible without making over complicated passages with a great number of links.

This tool is also expensive to run – it uses 4 credits for each classification of a Link (1 for each dichotomy), while a free account only gets 500 credits per day.

4. Evaluation

I compared the maximum scores achieved by the tool for each personality type (in Best Immediate Score mode) with the average score obtained by people of the same personality type.

The tool and the testers played the first level of "Creatures Such as We" by Lynnea Glasser [2]. I chose this game due to the diversity of its links' personality types.

I recreated the game in Twine format. The content of the original game was mostly left intact, with just a few suppressions of text after the player has made his last selection.

I focused the evaluation on the scores obtained instead of on the options selected because keeping track of several possible paths for each personality type would be unfeasible with the amount of personality types being analyzed, plus the scores give good approximations to the desirability of the selected options.

I set the Minimum Confidence Value to 0.501.

4.1. Theoretical Results

4.1.1. 4-Dichotomy Personality Types (16 possibilities)

Personality Type	Maximum Score
ESFJ	0.929
ISFJ	0.875
ISFP	0.875
ESFP	0.833
ENFJ	0.786
ISTJ	0.750
ESTJ	0.750
INFJ	0.708
ESTP	0.679
ISTP	0.667
INFP	0.667
INTP	0.667
ENTP	0.667
INTJ	0.625
ENFP	0.625
ENTJ	0.625

Table **6**: Maximum score obtained for each 4-dichotomy personality type

The maximum scores fluctuated between 0.625 (INTJ, ENFP, and ENTJ) and 0.929 (ESFJ), for an average of 0.733.

4.1.2. 3-Dichotomy Personality Types (32 possibilities)

Personality Type	Maximum Score
ESFx	0.952
xSFJ	0.952
ESxJ	0.952
ExFJ	0.952
xSFP	0.944
ISxJ	0.889
ISFx	0.889
ISxP	0.833
IxFP	0.833
IxFJ	0.833
ESxP	0.810
ExFP	0.778
xNFJ	0.778
xSTJ	0.778
ENxJ	0.762
INxP	0.722
IxTP	0.722
xNTP	0.722
ExTJ	0.722
ESTx	0.714
ENFx	0.714
INTx	0.667
ISTx	0.667
IxTJ	0.667
INFx	0.667
ExTP	0.667
INxJ	0.667
ENxP	0.667
xNTJ	0.667
xNFP	0.667
ENTx	0.667
xSTP	0.667

Table **7**: Maximum score obtained for each 3-dichotomy personality type

The maximum scores fluctuated between 0.667 (INTx, ISTx, IxTJ, INFx, ExTP, INxJ, ENxP, xNTJ, xNFP, ENTx, and xSTP) and 0.952 (ESFx, xSFJ, ESxJ and ExFJ), for an average of 0.768.

4.1.3. 2-Dichotomy Personality Types (24 possibilities)

Personality Type	Maximum Score
ExxJ	1.000
xSFx	1.000
xSxJ	1.000
ESxx	0.929
ExFx	0.929
xxFJ	0.929
xSxP	0.917
xxFP	0.917
IxxP	0.833
ISxx	0.833
IxFx	0.833
IxxJ	0.833
ExxP	0.786
xNTx	0.750
xNxJ	0.750
xNxP	0.750
xxTP	0.750
ExTx	0.714
INxx	0.667
IxTx	0.667
ENxx	0.667
xNFx	0.667
xSTx	0.667
xxTJ	0.667

Table **8**: Maximum score obtained for each 2-dichotomy personality type

As presented above, the maximum possible scores fluctuated between 0.667 (for INxx, IxTx, ENxx, xNFx, xSTx, and xxTJ) and 1.000 (for ExxJ, xSFx, and xSxJ), resulting in an average maximum score of 0.811.

4.1.4. 1-Dichotomy Personality Types (8 possibilities)

Personality Type	Maximum Score
Exxx	1.000
xSxx	1.000
xxFx	1.000
xxxJ	1.000
xxxP	1.000
xNxx	0.833
Ixxx	0.667
xxTx	0.667

Table 9: Maximum score obtained for each 1-dichotomy personality type

As presented above, the maximum possible scores fluctuated between 0.667 (Exxx, xSxx, xxFx, xxxJ, and xxxP) and 1.000 (Exxx, xSxx, xxFx, xxxJ and xxxP), resulting in an average maximum score of 0.896.

4.1.5. Discussion

As expected, as the criteria became more relaxed (i.e., the chances of hitting a maximum score increased from 1/16 to 1/2 by going from 4 dichotomies to 1), the average maximum scores increased as well.

Focusing only on the 4-dichotomy personality types, the story in question should be more enjoyable for people of the personality types with the highest maximum scores, with ESFJ (score = 0.929) being the best possible match, and less enjoyable for the personalities with the lowest maximum possible scores: INTJ, ENFP, and ENTJ (score = 0.625).

4.2. Practical Results

4.2.1. Test Format

4.2.1.1. Part 1 – Personality Test

The volunteers took an online personality test [3] to assess their Myers-Briggs personality type.

4.2.1.2. Part 2 – Data Collection and Gameplay

The volunteers were asked for the results of their personality test, their age group, and gender identity. After that, they played the Twine version of "Creatures Such as We."

The version used for the user featured a custom script that sent the user-provided information to a Google Sheets file. This way the volunteers had total anonymity and could complete the tasks at any time without supervision.

4.2.2. Data Analysis – Part 2: the Volunteers

I shared the personality test and game on Reddit, specifically on the /r/MBTI, /r/SampleSize, and the individual 16 personality types' subreddits (for example, /r/INTJ). 191 participants with valid answers took part in the study.

Age Group	Number of participants	Gender Identity		
		Male	Female	Other / prefer not to say
Under 18	46	18	28	0
18-25	92	47	28	7
26-35	35	18	16	1
36-45	12	2	8	2
46-60	4	1	1	2
60+	2	0	1	1
Total	191	86	92	13

Table **10**: Demographics of the participants by age group and gender identity

The following table contains the number of participants by personality type:

Personality Type	Number of Participants
INTP	29
INTJ	26
ISTP	22
ISTJ	21
INFP	14
ENFP	12
INFJ	11
ISFJ	11

ISFP	10
ENTJ	7
ESFP	7
ENTP	6
ESTP	5
ESTJ	4
ESFJ	4
ENFJ	2
Total	191

Table 11: Number of participants by personality type

9 personality types had at least 10 samples: INTP, INTJ, ISTP, ISTJ, INFP, ENFP, INFJ, ISFJ, and ISFP. For the rest of this document, a "relevant" personality type or dichotomy is one with at least 10 samples.

4.2.3. Data Analysis – Part 3: The Results

To analyze the volunteers' gameplays, I calculated the average scores of their gameplays relative to their personality type. After that, I averaged the results by personality type.

For each personality type, the "Average Human Score" column contains the average score achieved by people of that personality type, "Maximum Score" the maximum possible score, and the "Difference" the difference between the two.

4.2.3.1. 4-Dichotomy Personality Types (16 possibilities)

Personality Type	Average Human Score	Maximum Score	Difference
ESFJ (N=4)	0.667	0.929	0.262
ESTJ (N=4)	0.635	0.750	0.115
ISFJ (N=11)	0.628	0.875	0.247
ISTJ (N=21)	0.588	0.750	0.162
ENFJ (N=2)	0.583	0.786	0.203
ESFP (N=7)	0.583	0.833	0.250
INFJ (N=11)	0.564	0.708	0.144
ISTP (N=22)	0.560	0.667	0.107
ISFP (N=10)	0.542	0.875	0.333
ESTP (N=5)	0.529	0.679	0.150
INTJ (N=26)	0.527	0.625	0.098
ENFP (N=12)	0.480	0.625	0.145
INFP (N=14)	0.442	0.667	0.225
INTP (N=29)	0.412	0.667	0.255
ENTP (N=6)	0.396	0.667	0.271
ENTJ (N=7)	0.378	0.625	0.247

Table 12: Average human score and maximum score by 4-dichotomy personality type and the difference between the two.

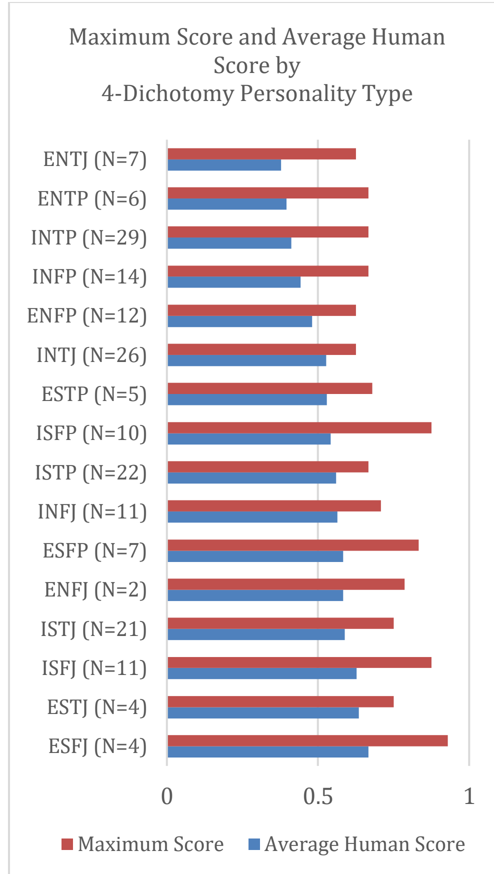


Figure 3: Maximum Score and Average Human Score by 4 -Dichotomy Personality type

The Average Human Scores fluctuated between 0.378 (ENTJ) and 0.667 (ESFJ), resulting in an average of 0.518.

The difference between the scores obtained by the tool and the average scores achieved by the volunteers of each personality type ranged between 0.098 (INTJ) and 0.271 (ISFP), for an average of 0.188, or 0.181 if only personality types with statistical relevance were considered. For the story in question, the tool was the most accurate at predicting INTJs' choices and the least accurate at predicting ISFPs', at least on a score-basis.

4.2.3.2. 3-Dichotomy Personality Types (32 possibilities)

Personality Type	Average Human Score	Maximum Score	Difference
ESxJ (N=8)	0.651	0.952	0.301
ExFJ (N=6)	0.639	0.952	0.313
xSFJ (N=15)	0.638	0.952	0.314
ESFx (N=11)	0.614	0.952	0.338
ISxJ (N=32)	0.602	0.889	0.287
IxFJ (N=22)	0.596	0.833	0.237
xSTJ (N=25)	0.596	0.778	0.182
ISFx (N=21)	0.587	0.889	0.302

ESTx (N=9)	0.576	0.714	0.138
ISTx (N=43)	0.574	0.667	0.093
xNFJ (N=13)	0.567	0.778	0.211
ESxP (N=12)	0.561	0.810	0.249
xSFP (N=17)	0.559	0.944	0.385
ISxP (N=32)	0.554	0.833	0.279
xSTP (N=27)	0.554	0.667	0.113
IxTJ (N=47)	0.554	0.667	0.113
INxJ (N=37)	0.538	0.667	0.129
ExFP (N=19)	0.518	0.778	0.260
INFx (N=25)	0.496	0.667	0.171
xNTJ (N=33)	0.495	0.667	0.172
ENFx (N=14)	0.494	0.714	0.220
IxFP (N=24)	0.484	0.833	0.349
IxTP (N=51)	0.476	0.722	0.246
ExTJ (N=11)	0.471	0.722	0.251
INTx (N=55)	0.466	0.667	0.201
xNFP (N=26)	0.459	0.667	0.208
ExTP (N=11)	0.456	0.667	0.211
ENxP (N=18)	0.452	0.667	0.215
ENxJ (N=9)	0.423	0.762	0.339
INxP (N=43)	0.422	0.722	0.300
xNTP (N=35)	0.410	0.722	0.312
ENTx (N=13)	0.386	0.667	0.281

Table **13**: Average human score and maximum score by 3-dichotomy personality type, and the difference between the two.

The Average Human Scores fluctuated between 0.386 (ENTx) and 0.651 (ESxJ), resulting in an average of 0.518, or 0.516 if only considering relevant cases.

The differences ranged between 0.093 (ISTx) and 0.385 (xSFP), for an average of 0.226, or 0.224 if only considering relevant cases.The tool was the most accurate at predicting ISTxs' choices and the least accurate at predicting xSFP s', at least on a score-basis.

4.2.3.3. **2-Dichotomy Personality Types (24 possibilities)**

Personality Type	Average Human Score	Maximum Score	Difference
xSxJ (N=40)	0.612	1.000	0.388
xxFJ (N=28)	0.605	0.929	0.324
ESxx (N=20)	0.597	0.929	0.332
xSFx (N=32)	0.596	1.000	0.404
ISxx (N=64)	0.578	0.833	0.255
xSTx (N=52)	0.574	0.667	0.093
IxxJ (N=69)	0.568	0.833	0.265
xSxP (N=44)	0.556	0.917	0.361
ExFx (N=25)	0.547	0.929	0.382
xxTJ (N=58)	0.538	0.667	0.129
IxFx (N=46)	0.538	0.833	0.295
ExxJ (N=17)	0.530	1.000	0.470
xNxJ (N=46)	0.516	0.750	0.234
IxTx (N=98)	0.514	0.667	0.153
xxFP (N=43)	0.499	0.917	0.418
xNFx (N=39)	0.495	0.667	0.172
ExxP (N=30)	0.495	0.786	0.291
IxxP (N=75)	0.478	0.833	0.355
INxx (N=80)	0.476	0.667	0.191
xxTP (N=62)	0.473	0.750	0.277
ExTx (N=22)	0.464	0.714	0.250
xNTx (N=68)	0.451	0.750	0.299
ENxx (N=27)	0.442	0.667	0.225
xNxP (N=61)	0.431	0.750	0.319

Table **14**: Average human score and maximum score by 2-dichotomy personality type and the difference between the two

The Average Human Scores fluctuated between 0.431 (xNxP) and 0.612 (xSxJ), resulting in an average of 0.518.

The differences ranged between 0.093 (xSTx) and 0.470 (ExxJ), for an average of 0.269. The tool was the most accurate at predicting xSTxs' choices and the least accurate at predicting ExxJs', at least on a score-basis.

4.2.3.4. **1-Dichotomy Personality Types (8 possibilities)**

Personality Type	Average Human Score	Maximum Score	Difference
xSxx (N=84)	0.583	1.000	0.417
xxxJ (N=86)	0.560	1.000	0.440
xxFx (N=71)	0.541	1.000	0.459
Ixxx (N=144)	0.521	0.667	0.146
Exxx (N=47)	0.508	1.000	0.492
xxTx (N=120)	0.504	0.667	0.163
xxxP (N=105)	0.483	1.000	0.517
xNxx (N=107)	0.467	0.833	0.366

Table 15: Average human score and maximum score by 1-dichotomy personality type and the difference between the two

The Average Human Scores fluctuated between 0.467 (xNxx) and 0.583 (xSxx), resulting in an average of 0.518.

The differences ranged between 0.146 (Ixxx) and 0.517 (xxxP), for an average of 0.344. The tool was the most accurate at predicting Ixxxs' choices and the least accurate at predicting xxxPs', at least on a score-basis.

4.2.4. **Discussion**

The Average Human Score remained consistent regardless of the number of dichotomies. As the numbers of dichotomies analyzed increased, the Average Maximum Score decreased. Consequently, the Average Difference values decreased as more dichotomies were considered.

Given the nature of the system, an Average Human Score just above the 50% threshold seems to indicate that the volunteers picked their links at random.

Though this might not mean anything in the great scheme of things, using the following formula, the Proportional Human Score values increase when more dichotomies are considered:

$$Proportional\ Human\ Score = \frac{Average\ Human\ Score}{Average\ Maximum\ Score}$$

Number of Dichotomies	Average Human Score	Average Maximum Score	Proportional Human Score
1	0.518	0.862	0.601
2	0.518	0.787	0.658
3	0.516	0.740	0.697
4	0.518	0.699	0.741

Table 16: Average human score, maximum score and proportional human score by number of dichotomies (only relevant cases)

By individual personality types, of the 9 relevant 4-dichotomy personality types, 6 had a Proportional Human Score above 0.700:

Personality Type	Proportional Human Score
INTJ (N=26)	0.843
ISTP (N=22)	0.840
INFJ (N=11)	0.797
ISTJ (N=21)	0.784
ENFP (N=12)	0.767
ISFJ (N=11)	0.718
INFP (N=14)	0.663
ISFP (N=10)	0.619
INTP (N=29)	0.618

Table 17: Proportional Human Score by 4-Dichotomy Personality Type

Of the 28 relevant 3-dichotomy personality types, 9 had a Proportional Human Score above 0.700:

Personality Type	Proportional Human Score
ISTx (N=43)	0.860
xSTP (N=27)	0.831
IxTJ (N=47)	0.831
INxJ (N=37)	0.807
xSTJ (N=25)	0.766
INFx (N=25)	0.744
xNTJ (N=33)	0.742
xNFJ (N=13)	0.729
IxFJ (N=22)	0.716
INTx (N=55)	0.699
ENFx (N=14)	0.693
ESxP (N=12)	0.692
xNFP (N=26)	0.689
ExTP (N=11)	0.684
ENxP (N=18)	0.677
ISxJ (N=32)	0.677
xSFJ (N=15)	0.671
ExFP (N=19)	0.666
ISxP (N=32)	0.665
ISFx (N=21)	0.660
IxTP (N=51)	0.659
ExTJ (N=11)	0.653
ESFx (N=11)	0.645
xSFP (N=17)	0.592
INxP (N=43)	0.585
IxFP (N=24)	0.581
ENTx (N=13)	0.579
xNTP (N=35)	0.567

Table **18**: Proportional Human Score by 3-Dichotomy Personality Type

Of the 24 relevant 2-dichotomy personality types, 5 had a Proportional Human Score above 0.700:

Personality Type	Proportional Human Score
xSTx (N=52)	0.861
xxTJ (N=58)	0.807
IxTx (N=98)	0.770
xNFx (N=39)	0.743
INxx (N=80)	0.713
ISxx (N=64)	0.694
xNxJ (N=46)	0.687
IxxJ (N=69)	0.681
ENxx (N=27)	0.663
xxFJ (N=28)	0.652
ExTx (N=22)	0.650
IxFx (N=46)	0.645
ESxx (N=20)	0.642
ExxP (N=30)	0.630
xxTP (N=62)	0.630
xSxJ (N=40)	0.612
xSxP (N=44)	0.606
xNTx (N=68)	0.601
xSFx (N=32)	0.596
ExFx (N=25)	0.589
xNxP (N=61)	0.574
IxxP (N=75)	0.574
xxFP (N=43)	0.544
ExxJ (N=17)	0.530

Table **19**: Proportional Human Score by 2-Dichotomy Personality Type

Of the 8 relevant 1-dichotomy personality types, 2 had a Proportional Human Score above 0.700:

Personality Type	Proportional Human Score
Ixxx (N=144)	0.781
xxTx (N=120)	0.756
xSxx (N=84)	0.583
xNxx (N=107)	0.561
xxxJ (N=86)	0.560
xxFx (N=71)	0.541
Exxx (N=47)	0.508
xxxP (N=105)	0.483

Table 20: Proportional Human Score by 1-Dichotomy Personality Type

The disparities amongst personality types with the same number of dichotomies, plus the Average Human Score being just above the 50% threshold are what make me skeptical about these results.

5. Conclusion

For this research project I developed a tool that simulates gameplays of interactive narrative games based on Myers-Briggs personality types.

From an interactive narrative game and one or more Myers-Briggs personality, the tool simulates how a person of that personality type would play the game i.e., what options they would select, based on the personality type of those options, obtained using text classifiers. After analyzing the story, the tool exports a PDF file with the best paths for the personality types previously selected.

If the game is not a branching narrative, the results of Best Immediate Score and Best-First Search will be the same and therefore only one of them appears in the report, but if it's a branching narrative, the report will show the results of the 2 modes.

For the story analyzed, the results of the user tests were mixed. The Proportional Human Scores seem to suggest a light tendency for people to select options close to their personality type. However, the average Human Scores' low values and sometimes staggering differences between the theoretical and the practical results seem to suggest there's no significance to the options picked by the volunteers.

The inconclusive results might be due to the method I used, the interactive narrative game, the personality test, the text classifiers, or the MBTI as a classification system may not even be appropriate to what I had in mind.

The tool would have to be tested with more stories and with more users to conclude whether it's possible to predict what choices people make based on their Myers-Briggs personality type when playing through an interactive narrative game.

5.1. Future Work

I believe the tool should become entirely browser based.

The tool should also support other evaluation systems, such as Sentiment Analysis, Horoscope - anything that can be extracted from text - while maintaining the same method.

The tool should also allow the user to select links in real time and proceed from there until the end of the gameplay. This would give the player better control over the tool and evaluate specific local paths instead of just the overall best paths.

I would also like to create my own text classifiers instead of relying on previously existing ones.

Another possibility would be to allow users to upload texts written from the point of view of the different personality types and train the classifiers on those texts. This could be especially interesting if the setting of the story wasn't the modern-day world.

REFERENCES

[1] Thue, David & Bulitko, Vadim & Spetch, Marcia & Wasylishen, Eric. (2007). Interactive Storytelling: A Player Modelling Approach.. 43-48.

[2] Glasser, Lynnea.: Creatures Such as We (2014), https://www.choiceofgames.com/creatures-such-as-we/, online; Retrieved October 31 2021

[3] Humanmetrics Inc.: Jung Typology Test (1998), http://www.humanmetrics.com/cgi-win/jtypes2.asp, online; Retrieved October 31 2021´Conference Name:ACM Woodstock conference

[4] prfekt.: Myers Briggs Lifestyle Function Classifier (2021), https://www.uclassify.com/browse/prfekt/myers-briggs-lifestyle, online; Retrieved October 31 2021

[5] J. Golbeck, C. Robles, M. Edmondson and K. Turner, "Predicting Personality from Twitter," 2011 IEEE Third International Conference on Priva-cy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, Boston, MA, 2011, pp. 149-156, doi: 10.1109/PASSAT/SocialCom.2011.33.

[6] Mairesse, Francois & Walker, Marilyn & Mehl, Matthias & Moore, Roger. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. J. Artif. Intell. Res. (JAIR). 30. 457-500. 10.1613/jair.2349.

[7] Klimas, C.: Twine - an open-source tool for telling interactive, nonlinear stories (2021), https://twinery.org/, online; Retrieved October 31 2021

[8] prfekt.: Myers Briggs Judging Function Classifier (2021), https://www.uclassify.com/browse/prfekt/myers-briggs-judging-function, online; Retrieved October 31 2021

[9] Fabio Celli, Bruno Lepri, Joan-Isaac Biel, Daniel Gatica-Perez, Giuseppe Riccardi, and Fabio Pianesi. 2014. The Workshop on Computational Personality Recognition 2014. In Proceedings of the 22nd ACM international conference on Multimedia (MM '14). Association for Computing Machinery, New York, NY, USA, 1245–1246

[10] Laws, R. 2001. Robin's laws of good GMing. Steve Jackson

[11] GamesDrachen, Anders & Canossa, Alessandro & Yannakakis, Georgios. (2009). Player modeling using self-organization in Tomb Raider: Under-world (Pre-print). CIG2009 - 2009 IEEE Symposium on Computational Intelligence and Games. 1 - 8. 10.1109/CIG.2009.5286500.

[12] Deng, Jeremiah & Kasabov, Nikola. (2000). ESOM: an algorithm to evolve self-organizing maps from online data streams. 6. 3 - 8 vol.6. 10.1109/IJCNN.2000.859364.

[13] Fernando de Mesentier Silva, Scott Lee, Julian Togelius, and Andy Nealen. 2017. AI-based playtesting of contemporary board games. In Proceedings of the 12th International Conference on the Foundations of Digital Games (FDG '17). Association for Computing Machinery, New York, NY, USA, Article 13, 1–10.

[14] Zook, Alexander & Fruchter, Eric & Riedl, Mark. (2019). Automatic Playtesting for Game Parameter Tuning via Active Learning.

[15] Holmgård, Christoffer, M. C. Green, Antonios Liapis and J. Togelius. "Automated Playtesting With Procedural Personas Through MCTS With Evolved Heuristics." IEEE Transactions on Games 11 (2019): 352-362.

[16] S. F. Gudmundsson et al., "Human-Like Playtesting with Deep Learning," 2018 IEEE Conference on Computational Intelligence and Games (CIG), Maastricht, 2018, pp. 1-8, doi: 10.1109/CIG.2018.8490442.

[17] Witten, Ian H. 2004. Text mining. 2004.

[18] prfekt.: Myers Briggs Perceiving Function Classifier (2021), https://www.uclassify.com/browse/prfekt/myers-briggs-perceiving-function, online; Retrieved October 31 2021

[19] prfekt.: Myers Briggs Attitude Function Classifier (2021), https://www.uclassify.com/browse/prfekt/myers-briggs-attitude/, online; Retrieved October 31 2021

[20] Agarwal, Apoorv & Xie, Boyi & Vovsha, Ilia & Rambow, Owen & Passonneau, Rebecca. (2011). Sentiment Analysis of Twitter Data. Proceedings of the Workshop on Languages in Social Media.

[21] M. N. Sahono et al., "Extrovert and Introvert Classification based on Myers-Briggs Type Indicator(MBTI) using Support Vector Machine (SVM)," 2020 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2020, pp. 572-577

[22] The Myers & Briggs Foundation.: MBTI Basics, https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/, online; Retrieved October 31 2021

[23] Cui, Brandon and Calvin Qi. "Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction." (2017).

[24] S. Bharadwaj, S. Sridhar, R. Choudhary and R. Srinath, "Persona Traits Identification based on Myers-Briggs Type Indicator(MBTI) - A Text Classification Approach," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 1076-1082

[25] Yannakakis, Georgios N., P. Spronck, D. Loiacono and E. André. "Player Modeling." Artificial and Computational Intelligence in Games (2013).