A Model of Character Evolution based on Stanislavsky-driven BDI Agents

Matthias Wilder\textsuperscript{1} and Pablo Gervás\textsuperscript{2}

\textsuperscript{1} University of Georgia
Athens, Georgia, USA
\textsuperscript{2} Instituto de Tecnología del Conocimiento - Facultad de Informática,
Universidad Complutense de Madrid
Ciudad Universitaria, 28040 Madrid, Spain
matthias.wilder25@uga.edu pgervas@ucm.es
WWW home page: http://nil.fdi.ucm.es/

Abstract. Many successful stories make use of changes in certain characters' beliefs, motivations, and relationships as a source of dramatic interest during the narrative. In this paper, we propose a character evolution model comprised of a character agent framework incorporating features of both the belief-desire-intention agent framework and Stanislavsky’s Method of Actions and a set of narrative events categories that produce characterization shifts. We describe how the elements of this model function and demonstrate the viability of the model to describe the character evolution of a number of characters in Shakespearean plays. Lastly, we discuss the ways the model might be extended and used in a story generation system to facilitate character evolution in the narrative.

1 Introduction

One element of a successful narrative is effective evolution of characters. While significant research has been conducted to explore how story generation systems can generate intentional [1], emotive [2], and even familiar [3] character agents, little research has focused on how such agents might change in characterization in a controlled form. This paper proposes a character evolution model to be used by a system to facilitate intentional development of agent characteristics.

Character evolution can be defined as any change to the perceived components of a reader’s mental model for a character [4]. For the purposes of this paper, we restrict our focus to evolution of a character’s fundamental beliefs and desires, rather than say their emotionality or physical appearance.

How characters evolve throughout a narrative has been a major focus of stories both popularly appealing and culturally potent. In the film series Star Wars, Darth Vader’s famous statement "I am your father" prompts a radical change in Luke’s understanding of his father’s past. This evolution of belief causes Luke to pity Darth Vader and to want to save him from darkness. In the ending of the original trilogy Luke asks Vader to fight against the Emperor, as opposed to attempting to kill Vader outright. Darth Vader ultimately sacrifices
himself for his son. This sequence of events features main characters moving from hating to sympathizing with enemies and from self-interest to self-sacrifice.

In this paper, we examine the evolution of a selection of characters from plays written by William Shakespeare in order to propose a computational framework that combines observed aspects of character evolution, an existing model for goal driven planning agents, and a theory for how motivations can inform an understanding of a character in a play.

2 Previous Work

Here we examine models of character used by past story generation systems and how they do not address the character evolution explicitly. Then we examine the features of the belief-desire-intention agent framework and Stanislavsky’s Method of Actions as inspirations for our proposed character agent framework.

2.1 Belief-desire-intention Framework

The belief-desire-intention (BDI) software framework is a base model for intelligent agents that relies on the separation of the knowledge (beliefs), goals (desires), and currently executing plans (intentions) of an agent [5]. It’s motivation for development has been the desire to find a middle ground between committing to certain plans of action and reconsidering others when necessary.

The formulation of beliefs, desires, and intentions in the BDI framework provides a possible analogy for character agents in a story world. An especially useful characteristic is the acknowledgement that beliefs are inherently subjective, i.e. beliefs represent knowledge that an agent assumes to be true but can be false. It also doesn’t specify that an agent must trust all beliefs completely, so an agent might acknowledge more or less certainty in different beliefs.

2.2 Stanislavsky’s Method of Actions

Stanislavsky’s Method of Actions is a learning tool that was developed for actors to learn how to conceive and portray their characters more believably through physical actions that expressed specific emotions.

It posits that at any point in the play a character has a specific objective, a goal towards which the character is working expressed as a verb phrase such as "to kiss her" [6]. Each objective is predicated on a motivation, defined abstractly in the method as the answer to the question: "Why does this character want this objective?".

Stanislavsky’s Method posits that character actions result from the character trying to resolve conflicts that hinders the achievement of their objective. Those actions produce emotional reactions, and the physical actions and emotional reactions can be related through an actor’s performance to the reader. Thus the method connects the high level concepts of motivations and objectives to the low level concepts of actions and reactions.
When a character’s objective changes, the situation that causes the change is named an event. These events can be seen as specific moments of character evolution, and justify the segmentation of a whole narrative into a series of character shifting events, a process that Stanislavsky called active analysis [6]. Such a segmentation provides a useful framework for computational systems to implement character evolution.

2.3 Existing Models of Characters in Story Generation

Previous story generation systems have handled characters in a variety of different ways. Many systems such as TALESPIN define planning agents to model characters with goals that could arise out of practical or emotional desires [1,3,7]. Some systems such as UNIVERSE [3] and TOK-EM [2] have focused on the personalization of characters, but the resulting remain relatively static over the course of the generated narratives.

MEXICA [8] includes methods for ensuring the development of characters’ emotional and tensional states over the course of a narrative. However, these developments such as getting angry with someone or becoming imperiled are limited to a character’s immediate context rather than persistent traits indicative of personalization.

Peñado, Cavazza, and Pizzi [9] conducted an investigation in a narrative BDI framework for representing characters in stories, but did not address how characters formulated following the model might evolve.

3 Methodology

We began our investigation by reviewing three Shakespearean plays. The Shakespearean works we draw from are Othello, Macbeth, and Much Ado About Nothing. We use these plays for the following reasons: the simplicity of linear plot progression, the diversity of narratives, the ability to separate the narratives into discrete events, and the way characters express internal cognition in explicit statements. We also deemed it more appropriate to manipulate the concepts from Stanislavsky’s Method of Actions in their original domain of plays.

From that investigation, we identified candidate characters that demonstrate interesting character evolution over the course of the narratives. By analyzing the features of their evolutions and the specific story events that they were involved in, we developed a framework for character agents and a number of character shift events in story generation systems that describe a model for character evolution in generated stories.

Finally we translated the narratives of the plays from the perspective of each character into a sequences of events including character shift events and simulated the effects of these sequences on a character agent using our proposed model. We examined the resulting representations of both our characters and their evolutions over the course of our narrative, comparing them to their original portrayals in Shakespeare’s works.

We discuss our findings and goals moving forward at the end of this paper.
4 Initial Explorations of Shakespearean Plays

For this investigation, we examined the evolution of the following characters: Macbeth and Lady Macbeth from Macbeth, Othello from Othello, and Benedict and Beatrice from Much Ado About Nothing.

Below we examine in depth the narratives of Macbeth and Othello, whose character evolutions are the most developed and detailed of the characters we examined.

In Macbeth, the character Macbeth is introduced undefined aside from some statements that he is well-respected. His interactions with the witches early in Act I produce beliefs in his future successes that drive him to murder Duncan quickly. Throughout the next several acts a combination of ambition and paranoia drive his desires and actions resulting in more murders. After every act of violence Macbeth demonstrates guilt which builds up. By the final act, Macbeth has confidence in his invincibility from the second round of prophecies, but also a deep seated pessimism about life and achievement that has grown from the many deaths his actions have caused. The traumatic experiences of the many deaths connected to Macbeth and the reassurance of the witches’ prophecies are the main sources of Macbeth’s beliefs and desires throughout the play. While he generally achieves his goals, the consequences of his actions tend to spoil his general mood and emotional state.

In Othello, much of the first two acts serve to set up the characterization of Othello. Othello’s character transformation takes place rather rapidly during Act III Scenes 3 and 4 through clever dialogue with Iago and then is reinforced and explored during the rest of Acts III and IV as he spies on and investigates Cassio and Desdemona. He undergoes another last minute evolution at the end of Act V wherein he realizes Desdemona was pure all along and he feels intense guilt and rage to the point where he believes that the only honorable action left to him is to commit suicide. Othello changes from honorable and levelheaded to jealous and violent due to Iago’s insinuations and manipulations of circumstance. This contrasts to Macbeth’s evolution which is based on less on persuasion and more on actions taken and events experienced. Othello’s evolving characterization is then related through his increasingly aggressive behavior to Desdemona and his willingness to believe Iago’s manipulations.

5 Character Evolution Model

Our model for how character evolution might work computationally has two main elements: an agent model with features for characterization and a set of story events that evolve characterizations. We present both in this section.
5.1 Character Agent Model

**BDI Framework** We begin by outlining the general structure of a BDI framework for character agents that adapts the definitions for the beliefs, desires, and intentions of an agent to the problem domain of story generation and character evolution.

Beliefs are a character agent’s concept of the story world. In the proposed framework a belief is defined as:

\[ B = (s, \delta, \gamma, C) \]

where \( s \) is a statement about the story world, \( \delta \) is an integer describing the disposition a character has towards \( s \), \( \gamma \) is as an integer describing the confidence a character has in \( s \), and \( C \) is a set of conditions, defined below, that would either affirm or negate \( s \).

A high positive value for \( \delta \) implies that a character agent derives intense positive emotions from \( s \) and the opposite is true of a low negative value. A value of zero for \( \delta \) implies that character has no emotional relationship with \( s \).

A high value for \( \gamma \) implies that a character agent has a lot of confidence in \( s \). A low value or value of 0 for \( \gamma \) implies that a character has little or no confidence in \( s \). In this paper, we utilize the range -8 to 8 for the value of \( \delta \) and 0 to 8 for \( \gamma \). Beyond requiring that \( \gamma \) be non-negative and that \( \delta \) range between two values of opposite sign but equal magnitude, these restraints are arbitrary.

Each element of \( C \) is a condition, \( c = (o, i) \), where \( o \) is a statement about the story world and \( i \) is an integer describing the impact that \( c \) has on \( \gamma \). In simple terms, if a character observes that \( o \) is true then \( \gamma \) is either increased or decreased by the value of \( i \). Thus if \( i \) is positive then \( c \) is called an **affirm condition** for \( B \) because observing \( o \) would increase \( \gamma \), i.e. a character’s confidence in \( s \). If \( i \) is negative then \( c \) is called an **negate condition** for \( B \) because observing \( o \) would decrease \( \gamma \).

As an example, we declare "Desdemona loves me" an important belief for the character Othello. As a statement for the story world, we translate that belief as the statement \( \text{loves}(\text{Desdemona}, \text{Othello}) \) for \( s \). At the beginning of his play, Othello more or less completely believes this statement and views it positively, therefore we set \( \delta \) and \( \gamma \) to be positive values +5 and +6, respectively. An observation that decreases Othello’s confidence in this belief is that Cassio and Desdemona had sex so we add to \( C \) a negate condition \( (\text{had-sex}(\text{Desdemona}, \text{Cassio}), -4) \).

Desires are goals characters wish to pursue, either in order to affirm a positive belief or negate a negative belief. They are defined in this framework as

\[ D = (o, i, B) \]

where \( o \) is a statement about story world and \( i \) is an integer describing the magnitude of the impact that the achievement of this desire will have on the confidence of \( B \), the belief being targeted.
Desires are formed automatically from a condition, $c$, of $B$ when $B$ is generated, with $o$ and $i$ being taken directly from the definition of $c$. If $B$ is a positive belief ($\delta > 0$) then the $c$ chosen will be an affirm condition. If $B$ is a negative belief ($\delta < 0$) then the $c$ chosen will be an negate condition.

Intentions are plans currently in execution. Intentions are comprised only of a plan whose definition is outside the scope of this paper.

**Stanislavsky’s Method Components** In order to extend the BDI framework to facilitate character evolution, here we propose added features for our agent model that describe certain elements of Stanislavsky’s Method of Actions.

The identification of emotional states of a character in a narrative is an important element of Stanislavsky’s Method. For the purposes of this paper, we model a character agent’s general emotional state as an integer value, $\phi$, we name the agent’s affect. A positive value for $\phi$ implies the character on the whole feels happy, content, confident, etc. and a negative value implies the character feels sad, angry, concerned, etc.

At any point in the story we calculate $\phi$ for a character agent holding a set of beliefs $B_1, ..., B_n$ where $B_i = (s_i, \delta_i, \gamma_i, C_i)$ as the sum of the products of each belief’s disposition and confidence values:

$$\phi = \sum_{i=1}^{n} \delta_i \gamma_i$$

Accordingly, high confidence in highly positive beliefs and low confidence in highly negative beliefs generate a positive emotional affect. The opposite conditions generate a negative emotional affect.

Furthermore, when confidence in a positive belief is increased a positive emotional response is triggered in the character agent i.e. $\phi$ is increased as well. When confidence in a positive belief is lowered a negative emotional response is triggered in the character agent. The effects are reversed for negative beliefs and are non-existent for neutral beliefs ($\delta = 0$).

This feature of a character agent allows for underlying changes in beliefs which may not be expressed directly to a reader to be expressed indirectly in the emotions a character agents exhibits in the text.

Crucially, following in Stanislavsky’s Method, a character agent must select among their current desires one desire to be turned into an objective. Alternatively, another character may convince a character to take up a specific objective as long as it satisfies a desire and does not conflict with other desires. By explicitly connecting the definition of a desire to a specific belief, we ensure that every objective has a motivation implicitly defined.

Once an objective is selected, a plan can be made to achieve it. This plan, or intention in the BDI framework, directs the character agents actions until an
Thus our character agent model is able to mimic Stanislavsky’s conceptualization of character behavior wherein a character with certain motivations forms an objective and attempts to achieve that objective by performing actions whose consequences produce emotional responses and potentially a change in objectives. As the underlying beliefs and desires of characters are evolved, corresponding changes in the character’s affect and actions taken become evident at the textual level, effectively demonstrating character evolution to a reader.

5.2 Character Shift Events

In order for character evolution to be present in a story using the above character agent model, beliefs (and by extension desires) must change as a result of events in the narrative. Here we propose four distinct categories of events that produce changes in an agent’s characterization.

**Traumas and Miracles** Traumas and miracles are story events that generate new negative and positive beliefs, respectively. Unless a character begins the narrative with a number of beliefs set, as in the case of Othello, traumas and miracles will drive the characterization of an agent, because the resulting beliefs will dictate the initial desires, objective, and aspect of a character agent.

There are a wide variety of trauma and miracle events in the plays we’ve examined, but an example of a miracle event we identified is Macbeth hearing the first of the witches prophecies:

\[
X \text{ hears-first-prophecies}
\]

\[
\text{add-belief}(X, "I'm going to become king")
\]

\[
\text{add-belief}(X, "Banquo's children will become kings")
\]

Hearing the original prophecies from the witches produces in Macbeth two important beliefs: "I'm going to become king" which we determine has a disposition value of +5 and a confidence value of 2 at its inception and "Banquo’s children will become kings" which we determine has a disposition value of -2 and a confidence value of 2 at its inception. We also determine that they both contain \((\text{thane-of-cawdor}(\text{Macbeth}), +4)\) among their affirm conditions, which becomes clear once Macbeth is named Thane of Cawdor and Macbeth immediately reacts by declaring his reinforced belief that he will be king, and by extension Banquo’s children will also be kings.

Note that although the belief "Banquo’s children will be kings" has a negative disposition, its effect on Macbeth’s emotional affect is overpowered by the positive effect generated by the belief "I'm going to become king". Thus this event is classified as a miracle.

hears-first-prophecies is very specific to the plot of Macbeth, however some tragedies and miracles we determine to be generalizable as in the case of
X kills Y
  add-belief(X, "Y is dead")
  add-belief(X, "I killed Y")
  add-belief(X, "I am a murderer")

which functions similarly across the different plays we examine. When we defined
the specific beliefs added by this event, we separated an emotional belief, "I am
a murderer" which we determined has a highly negative disposition value, from
descriptive beliefs "Y is dead" and "I killed Y" which we determined have
disposition values of zero and thus are neutral statement about the story world.

It should be mentioned, though, that neutral beliefs from traumas and mira-
cles might change the character’s emotional affect if they trigger affirm or negate
conditions for other beliefs.

Revelations Revelations are events wherein a character agent is made aware
of a neutral already existing aspect of the story world.

Revelations are distinct from traumas and miracles in that a revelation event
do not represent a change in the story world, but rather an expansion of a
carer’s understanding of the current state of the story world. For instance,
the trauma X kills Y implies that Y was alive but now Y is dead. The revelation
X learns dead(Y) implies that Y was dead already and X just didn’t know it. So
for a general statement about the world, o, we identified the following revelation
events:

  X learns o
  add-belief(X, "o is true")

  X realizes o
  add-belief(X, "o is true")

Y tells X o
  add-belief(X, "o is true")

where the belief "o is true" is neutral, is completely believed, and only has a
negate condition for the statement ¬o, in case there is a new development or
revelation later on in the plot.

A concrete instance of this type of event in Othello is when Othello learns
that Emilia gave Desdemona’s handkerchief to Iago and Cassio never really had
it.

It is important to note that revelations need not necessarily be true in terms
of the real state of the story world. Something being learned or told may be
false. Beliefs for a character agent reflect the assumptions that the agent has
about the story world, and at times it is important that a character assumes an
incorrect fact for greater dramatic effect later on down the road.

Commitments Commitments mark moments when a character agent selects
an objective from their desires. The most simple commitment event is
X decides D
set-objective(X,D)

which describes character X independently deciding to pursue the desire D. D would be chosen among the desires based on how powerfully it could impact X’s emotional affect.

An alternative commitment event we identified is

Y convinces X D
set-objective(X,D)

which describes character X deciding to pursue core desire D per the influence of character Y. This case would require Y to have as part of their plan to make X achieve D. It would still require D to be among X’s desires, though.

For each case, the current objective of character agent X is set to D, and a plan to achieve it is added to X’s intentions.

Resolutions Resolutions describe either the success, failure, or abandonment of a character agent’s objective. Accordingly there are only three template events:

X succeeds D
remove-objective(X,D)

X fails D
process-failure(X,D)
remove-objective(X,D)

X abandons D
remove-objective(X,D)

A succeed event is chosen when the character believes the goal of the objective is met. Confidence in the underlying belief for the objective will have already increased or decreased accordingly when the goal objective was absorbed into the character’s knowledge of the world (incidental beliefs).

A fail event happens when the character believes the next event in a core intention is no longer executable. In this case, confidence in underlying belief is increased or decreased as an opposite effect to what the character was hoping to achieve. This is implemented via the process-failure part of the event definition. If a character was trying to affirm a belief then confidence is decreased. If a character was trying to negate a belief then confidence is increased.

An abandon event happens when an objective is no longer important enough to be pursued or if another desire becomes important enough to replace the current objective. Nothing happens to the confidence in the underlying belief when a character abandons an objective.

For each case, the current objective and its corresponding intention for the character agent are removed and a new commitment event must occur to give the character a new objective.
6 Othello: A Small Walkthrough

Here we relate a small section of the translation of Othello’s narrative into our model for character evolution.

We start from the moment Iago tells him that he suspects Iago and Desdemona are having an affair. At this point in the play we determine that Othello has the following relevant core beliefs:

"Desdemona loves me" = (loves(Desdemona, Othello), +5, +6, C_1)
"Cassio is a good person" = (good-person(Cassio), +3, +6, C_2)

For the sake of simplicity we will only provide relevant affirm and negate conditions for each core belief discussed in this section. As such we have the only relevant conditions being

\[ C_1, C_2 \supseteq \{ (\text{had-sex(Desdemona, Cassio)}, -4) \} \]

Also, currently Othello has no objective since he just finished his objective of punishing Cassio by stripping him of his rank.

Because Iago desires to trick Othello, the next event is

Iago implies-affair-between (Cassio, Desdemona)

which is a trauma that inspires the following core beliefs in Othello:

"Desdemona and Cassio are in an affair" = (affair(Desdemona, Cassio), -5, +2, C_3)
"Desdemona needs to be punished" = (guilty(Desdemona), -3, +2, C_4)
"Cassio needs to be punished" = (guilty(Cassio), -3, +2, C_5)

with some overlapping conditions:

\[ C_3, C_4, C_5 \supseteq \{ (\text{had-sex(Desdemona, Cassio)}, +7), (\text{has(Cassio, handkerchief)}, +2), (\text{negated-had-sex(Desdemona, Cassio)}, -7), (\text{has(Desdemona, handkerchief)}, -2) \} \]

and some non-overlapping conditions:

\[ C_4 \supseteq \{ (\text{dead(Desdemona)}, -7) \} \]
\[ C_5 \supseteq \{ (\text{dead(Cassio)}), -7 \} \]

The impact of this event is that Othello begins to believe that Cassio and Desdemona have betrayed him, but he is not completely sure. This doubt is reflected by the fact that his confidence value for each of these new beliefs is only 2. Even so, his affect value, \( \phi \), has been decremented by 17 points to these new negative beliefs due to their negative disposition values.

The two beliefs "Desdemona needs to be punished" and "Cassio needs to be punished" reflect a desire in Othello to kill both of them for their betrayals.

Now Othello needs to take on an objective and thus

Othello decides to-exonerate(Desdemona)

which makes Othello’s objective to-exonerate(Desdemona) which uses motivation of negate the belief"Desdemona and Cassio are in an affair". The objectives goal is \( \neg \text{has(Desdemona, handkerchief)} \), a negate condition for the
belief. Thus Othello goes to indirectly confront Desdemona about the handkerchief whereupon

Othello learns $\neg$ has(Desdemona, handkerchief)

and thus

Othello fails to-exonerate(Desdemona)

since Othello does not have any more clear paths to exonerating Desdemona. Because he failed to negate the belief, Othello’s confidence in "Desdemona and Cassio" are having an affair is increased, and thus his affect becomes more negative.

Once again, Othello has to find a new objective so

Othello decides to-exonerate(Cassio)

So he goes and spies on Cassio. Othello overhears Cassio and Iago talking and

Othello learns had-sex(Cassio, Desdemona)

Othello learns has(Cassio, handkerchief)

so

Othello fails to-exonerate(Cassio)

All of which dramatically increases Othello’s belief in "Desdemona and Cassio are in an affair" as well as "Desdemona needs to be punished" and "Cassio needs to be punished". Without any possibility for negating "Desdemona and Cassio are in an affair" though, Othello at this point gives up on trying to disprove the affair and thus changes his motivation to negating another belief, so that his new commitment becomes

Othello decides to-kill(Desdemona)

At this point Othello has evolved with some very strong negative beliefs and a very negative emotional affect. Now, to satisfy his belief that Desdemona should be punished, he aims to kill his former love, an objective which precipitates the tragedy that ends the play.

7 Discussion

Here we discuss aspects of the translation process, some limitations discovered during translation, and some directions for future investigation.

At this point, the translation process is performed manually and thus it is difficult to evaluate the accuracy of the resulting representation of the narratives. This issue is further complicated by the lack of an empirical comparison. That being said, most major inflection points for each of the characters’ narrative and evolution fit into the model relatively easily and we were able to mirror the general changes in their emotional states and goals through the character agent framework’s affect and objective. Figure 1 shows the evolution of emotional affect for each character considered according to our translation of the plays.
Fig. 1. Change in emotional affect for each character during their involvement in the narrative. Story time denotes a relative progression of actions from the "beginning" of a character's story to the "end".

The most difficult element of the character narratives to translate was the objective. At times more than one objective seemed to control character’s actions, such as when Benedick offers his help to Leonato, presumably to help Hero, and then confesses his love to Beatrice, which does not directly help Hero. One objective might not be enough to accurately describe the behaviors of more complex characters with competing interests.

Another feature of our characters’ evolutions not effectively captured in this model at this stage are the complex logic among beliefs. For instance, Othello does not really understand the significance of the handkerchief to the idea of an affair between Desdemona and Cassio until Iago plants the idea into his mind that it is a sign. These changes to belief disposition and conditions after belief generation are captured by our model.

Nevertheless, we believe the system proposed in this paper merits more extensive investigation including integration with more complex models of character like those developed by Reilly and Bates[2] and Turner[10] and development from the analysis of more diverse story examples outside the plays discussed.

We eventually hope to develop algorithms for how character and authorial agents might interact with our model. The advantages of revision conscious story generation systems like the one developed by Perez [8] might also help handle the complexity of ensuring both character evolution and story cohesion. As a final goal, we hope to build a story generation system to examine the kinds of evolution that our model can generate on its own.
8 Conclusion

We set out to develop a model of character evolution that could be incorporated into a story generation system. Our resulting model combined elements from the BDI agent framework and Stanislavsky’s Method of Actions in conjunction with observed features of a selection of Shakespearean characters’ narratives. We translated these characters’ narratives into a sequence of character shift events that fit our model. These sequences represented the evolution of these characters as captured by our model. Finally, we discussed the limitations of our model given the resulting translations and proposed future areas of investigation for our model.

References