Australian Government
Department of Defence
Science and Technology

Event Sequencing for Situation Narratives

Ian Dall and Bradley Donnelly

Joint and Operations Analysis Division
Defence Science and Technology Group

DST-Group–TR–3351

ABSTRACT

Situation Awareness is a critical factor for decision making in complex dynamic environments. Information fusion techniques enable machines to augment human situation awareness, provided the machine representation of the situation can be conveyed to the human operator reliably.

The problem, addressed in this report, is to order a set of events (a plot) in such a way as to result in a narrative which maximises audience engagement, but does not mislead with respect to the temporal relationship of significant events. A novel algorithm is proposed which uses a contrained optimisation approach to achieve an efficient implementation. Measures of *temporal tension*, *continuity* and *centrality* are proposed and used as optimization criteria. Applying different weights to these measures produces narratives with different characteristics.

The algorithm has been implemented and the results of testing using a large historical naval battle (the World War II sinking of the battleship Bismarck) are presented. With appropriate weights the algorithm was shown to produce a plot comparable to human authored text. A number of vignettes are analysed which provides insight into the operation and limitations of the algorithm. In particular, incompleteness in the input data permitted misleading output in some cases. While unsurprising, this highlights the need for careful consideration of the affect of incompleteness and uncertainty on automatic situation assessment aids.

The need for future work has been identified including: refinement of the optimisation measures; broader and more rigorous evaluation; and integration into a complete situation assessment system.

RELEASE LIMITATION

Published by

Joint and Operations Analysis Division
Defence Science and Technology Group
506 Lorimer St,
Fishermans Bend, Victoria 3207, Australia

Telephone: 1300 333 362
Facsimile: (03) 9626 7999

AR-016-821
April, 2017

Event Sequencing for Situation Narratives

Executive Summary

Situation Awareness is a critical factor for decision making in complex dynamic environments. Information fusion techniques enable machines to augment human situation awareness, provided the machine representation of the situation can be conveyed to the human operator reliably.

Situations generally evolve over time and situation awareness requires the ability to project this into the future so as to assess likely impacts. In complex situations involving multiple actors, telling events in strict time order tends to result in a disjoint narrative. Audience engagement may be enhanced by following one actor for a some time before switching to follow another actor. The problem, addressed in this report, is to find an order of events (a plot) which maximises audience engagement but does not mislead with respect to the temporal relationship of significant events.

This report proposes a Multiple Intersecting Chronologies (MIC) model, inspired by the structure of fictional narratives such as *The Lord of the Rings* [Tolkien, 1966], which is flexible enough to represent narratives aimed enhancing situation awareness, but sufficiently constrained to allow efficient implementation. In this model each actor has its own strictly chronological story line, but story lines are asynchronous with respect to each other. Story lines may intersect or merge at which point they become synchronised.

A novel algorithm has been developed which takes an unordered list of events as input, constructs multiple story lines and traverses them in a way which respects the constraint that each story line be strictly chronological. Measures of *temporal tension*, *continuity* and *centrality* are used as optimization criteria. Temporal tension refers to the degree to which events are presented out of order, continuity is the degree to which the narrative continues to follow the events experienced by one entity and centrality was introduced to allow greater emphasis on following the story line of central characters. Applying different weights to these measures produces narratives with different characteristics. High temporal tension weight produces a narrative which tends towards a strict chronology. High continuity weight tends to produce a narrative which follows the action from cause to effect. High centrality weight tends to produce a narrative which follows the central character and defers telling the background of secondary characters until they interact with a more central character.

The algorithm has been implemented and was tested using a large historical naval battle (the World War II sinking of the battleship Bismarck). With appropriate weights the algorithm was shown to produce a plot comparable to human authored text.

A number of vignettes are analysed which provides insight into the operation and limitations of the algorithm. In particular, incompleteness in the input data permitted misleading output in some cases. While unsurprising, this highlights the need for careful consideration of the affect of incompleteness and uncertainty on automatic situation assessment aids.

The testing of the algorithm has yielded significant insight into its characteristics and limitations, however it can not be considered a rigorous evaluation. The output was evaluated

by comparing statistics with human authored text, but this leaves unanswered the question of whether these are the relevant statistics. Also, human authored text is not necessarily the 'gold standard'. Depending on the case, it may well be possible for the machine generated event sequence to be better than that found in human authored text. It would be more valid to assess audience attention, recall and comprehension or their ability to perform a relevant task although such methods are expensive and often impractical [Reiter and Belz, 2009]. Future work should include this kind of evaluation.

Only one scenario (albeit a large one) has been used leaving open the question of whether the algorithm is generally applicable and whether the weights need to be tuned for each scenario. In particular *temporal tension* has been identified as being problematic because of the varying tempo within scenarios and between scenarios. There is considerable scope for future work to investigate better measures of the *temporal tension* or replacing it with a new concept.

Finally, event sequencing is just one part of the automated narrative generation problem. Integration of this work into a complete system is necessary to achieve the goal of enhanced situation awareness. This would also enable rigorous human evaluation as discussed above.

Authors

Ian Dall
Joint and Operations Analysis Division

Ian Dall graduated with a Bachelor of Engineering (Honours) from the University of Queensland in 1978, and subsequently completed a Doctor of Philosophy degree in the area of non-linear signal processing, at the University of Adelaide in 1991. He joined Radar Division of the then Defence Research Centre Salisbury in 1983 to work on the Jindalee OTHR. Ian has worked on electronic design, automated fault detection, signal processing, data fusion, expert systems, dynamic games, decision aids and immersive visualisation. He currently leads Situation Assessment group, concerned with discovery, access, dissemination, interpretation and presentation of information for enhancing operational situation awareness.

Bradley Donnelly
Joint and Operations Analysis Division

In 2013 Bradley Donnelly was a participant of the Summer Vacation Placement(SVP) program at DST group. This piqued his interest in research and after he completed his Bachelor of Engineering(Robotics) and Master of Engineering(Electronics) in 2014 at Flinders University of South Australia, he begun his PhD with Flinders University and the ARC Research Training Centre for Naval Design and Manufacture. His PhD is on the acoustic properties of marine antifouling technologies and marine biofouling. He still has interests in instrumentation, signal processing, electronic design and AI.

THIS PAGE IS INTENTIONALLY BLANK

Contents

THIS PAGE IS INTENTIONALLY BLANK

1 Introduction

Situation Awareness is a critical factor for decision making in complex dynamic environments. One definition of situation awareness [Endsley, 1988] is 'the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.' An Information Fusion system can be used to augment the situation awareness of a human decision maker by automatically fusing information from multiple sources, identifying and analysing trends and anomalies for a particular situation, and predicting what the impact may be on the decision makers' goals. However, key to achieving this is that the machine representation of a situation, in the form of sets of statements in a formal language, needs to be conveyed reliably to the decision maker. One approach being explored is the use of multimedia to convey contextually rich representations of situations from a fusion system to decision makers [Wark and Lambert, 2007]. This is not the only possible approach and in other work (unpublished) users are able to explore the situation by means of interactive search.

Within the paridigm of multimedia presentation, this report focuses on the narrative mode of discourse. There is no universally agreed definition of narrative [Sternberg, 2010], but for the purposes of this report we will adopt the following definitions.

> A *narrative text* is a text in which an agent or subject conveys to an addressee . . . a story in a particular medium. . . A *story* is . . . the fabula presented in a certain manner. A *fabula* is a series of logically and chronologically connected events that are caused or experienced by actors. . . An *event* is the transition from one state to another state. — Bal [2009]

Bal uses the term 'text' in a very broad sense which includes cinema and visual art. We will prefer the term *presentation* reserving 'text' to mean a sequence of written symbols. And where Bal uses the term 'story' we will instead use the term *plot*. Consistent with these definitions, *plot generation* — the process of producing the plot — may include synthesising the fabula, selecting a subset of events and determining the order in which those events are to be told.

Other modes of discourse are possible: Besides narrative, Genung [1900] identifies description, exposition and argument as modes of discourse (Genung uses the term 'literary types'). We choose to focus on narrative because we anticipate content in which events predominate, such as the story of the sinking of the battleship Bismarck during World War II. Also as noted by Smith [2003], narratives 'often play a supporting role in essays, newspaper articles and other discursive genres.'

Wark and Lambert [2007] propose an Intelligent Multi-Media Presentation (IMMP) system which is responsible for translating the machine statements and presenting them in a way which most effectively communicates with human decision makers. The IMMP system needs to generate the *plot* and the *presentation* of the narrative in a way most appropriate for the content and the audience.

A system for augmenting situation assessment incorporating Information Fusion and IMMP sub-systems is under development (Figure 1). The separation between the fusion system and the IMMP system is characterised by the kind of knowledge utilised. The fusion system reasons about observed behaviour and intentions, taking into account knowledge of the entities, behaviour, the environment, etc. The IMMP system reasons about the best way to present content taking into account knowledge of the operator's situation awareness needs, the characteristic of the media and the art of story telling. There are significant challenges to implementing each component of the augmented situation assessment system and progress is incremental. The scope of the work described in this report is confined to event ordering which is one aspect of multi-media planning (or *discourse planning* in natural language generation terms [Reiter and Dale, 1997]).

Although our eventual application is multi-media presentation, this report addresses an aspect of planning which is independent of the medium and therefore also part of natural language generation. In natural language generation systems, Reiter and Dale [1997] identify six stages of processing:

1. content determination
2. discourse planning
3. sentence aggregation
4. lexicalization
5. referring expression generation
6. linguistic realisation

For IMMP, we identify four functions which must be performed by the IMMP (Figure 1) which subsume Reiter and Dale's six stages of processing:

1. content selection
2. planning
3. media assignment
4. finishing

Figure 1 shows only the main path by which content is transformed into a presentation. In practice, there may need to be a mechanism whereby, for example, content selection can be influenced by the media assignment.

Content selection is the process of selecting the relevant information to be presented which corresponds to the content generation step identified by Reiter and Dale [1997]. Content selection could be driven by user interrogation — for example: 'How was the Bismarck sunk?' Or significant events can be identified by reasoning about user's goals, role and state of

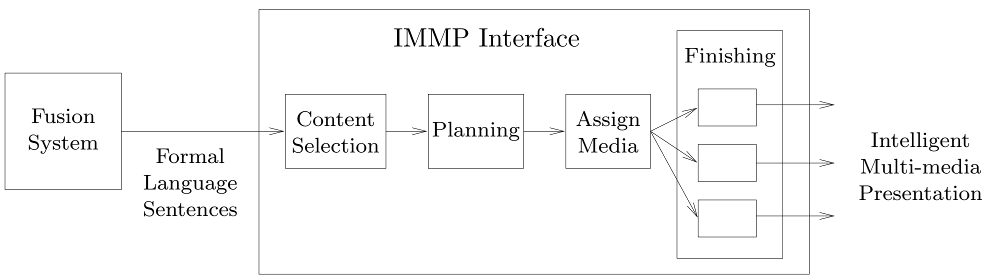

Figure 1: Augmented Situation Assessment

knowledge. Events which may impact on the user's goals would be deemed significant[1]. In any case, once significant events have been identifed related content may be selected by constructing a dependency tree of events.

Media assignment is the process of allocating which content is best presented by which media. For example the geo-spatial relationship between entities (including passive entities like boundaries and land masses) may be best presented graphically on a map, whereas intentions may be better presented as written or spoken text. Media assignment is not necessary in the case of natural language generation where there is only one medium.

Finishing is a generic term to denote the media specific processing. In the case of natural language generation finishing encompasses the steps of sentence aggregation, lexicalization, referring expression generation and linguistic realisation identified by Reiter and Dale [1997].

This report describes the design, implementation and evaluation of an event sequencing algorithm using the concepts of *continuity*, *temporal tension*, and *centrality* to order the events to be presented in a narrative. For evaluation purposes, this approach has been applied to an historical scenario, the sinking of the Bismarck during World War II, as it incorporates multiple actors within a militarily relevant scenario, with a wealth of historical analysis available.

In a multi-media presentation, there is the potential to present multiple story lines in parallel. However, the attention of the viewer is still limited and presenting a single narrative thread is a demonstrably effective approach. This simplification allows the planning function to be performed independently of the media assignment.

We assume in this report that relevant content has been selected and the planning function is to produce a single narrative thread conveying the selected content. Generation of multiple narrative threads could be considered as an extension to the current work. Discourse planning in full includes aspects such as context setting, elaboration, summarization and explanation. We assume, in this report, that determining the *order* of events to be related is separable from determining *how* to relate those events.

2 Related Work

An early example of automated plot generation is TALE-SPIN [Meehan, 1977], which simulates a small world of characters presented with problems which they must react to (or solve). This was coupled with an English generator. Notably there was no attempt to improve plot quality distinctly from creating the fabula — the events were told as they occurred in the simulation without selection or re-ordering.

More recent work has explored a variety of approaches. Gervas et al. [2005] used case-based reasoning to construct new fairy tales using a case database, constructed from traditional fairy tales and an ontology which defines permissible relations. Bui, Abbbass and Bender

[1] In estimating significance, various levels of sophistication are possible from filtering events based on value of assets and region of operations, through to complex reasoning about possible impact on current plans taking into account an adversary's capability and intention.

[2010] used a genetic algorithm to evolve new plots from a fabula derived from the *Little Red Riding Hood* traditional folk story. Permissible changes were defined using a regular grammar. Notably the evolution used human evaluation to determine fitness. Kybartas and Verbrugge [2014] used a graph re-writing technique, which they view as a generalization of grammar based techniques, to generate computer game narratives. These approaches all address the problem of creating fictional narratives. In other work Bui, Bender and Abbass [2012] addressed the problem of generating and narrating future scenarios. While future scenarios may be required to accurately convey some objective assessment of the most likely or most dangerous possible future, these scenarios [Bui, Bender and Abbass, 2012] are more akin to fictional narrative in so far as novelty was a goal. Fictional narratives have the goal of producing novel, plausible and interesting stories whereas this report is concerned with accurately conveying awareness of real world events effectively to the audience.

Automated plot generation for non-fiction[2] appears to have received little attention. Weather forecasting is a successful application of natural language generation which has a narrative aspect: It includes actors (abstractions such as cold fronts), events and causal relationships although these actors do not have intentions. Sripada et al. [2014] report on a system currently in use at the UK Met Office, however it produces short localised reports which do not require much in the way of plot generation. Sripada et al. [2014] do not give much detail of the system, citing commercial sensitivity, but earlier related work focussed on content generation [Sripada et al., 2001] and lexical choice [Reiter and Sripada, 2002]. Portet et al. [2009] note: 'Many of these systems have fairly simple document planners, while the nature of the data affords quite simple solutions for microplanning.'

BabyTalk [Portet et al., 2009], is a system for generating textual summaries from neonatal intensive care data. It operates in a more complex domain than weather forecast summaries, needing to deal with physiological data in the form of time series as well as discrete events such as medical interventions, human observations and laboratory test results. In BabyTalk, plot generation is performed by the document planning process which produces a sequence of key events. Key events, together with events which are explicitly linked to the key events, or co-temporal to them, are grouped in the same paragraph. Portet et al. do not state how the paragraphs are ordered. Since the plot may contain events out of chronological order, the microplanning process must generate linguistic marks to maintain temporal coherence: Paragraphs start with an explicit mention of the start time and changes of tense are used to indicate relative time. Even with these measures, time was not always communicated well and human generated texts had much better narrative structure than those generated by BabyTalk. Portet et al. [2009] note that 'temporal reasoning and representation [...] has received very little attention from the [natural language] generation point of view.'

In common with BabyTalk, our work on augmentating situation awareness is also concerned with narrative structure and with communicating temporal relationships. BabyTalk can also be considered a means to augment situation awareness, however the kinds of situations we consider differ. In particular we consider situations involving multiple independent actors with different and often conflicting goals, which is in this respect similar to many fictional narratives.

Plot generation for fiction is a problem with different characteristics than plot generation

[2] The term *plot* is not widely used in the context of non-fiction, however our definition of plot is equally applicable to fiction and non-fiction.

for non-fiction. In general, a fictional narrative may be constructed from the entire set of plausible actors, objects and events which may be sequenced in any plausible order. Although what is considered plausible varies by genre, the plot solution space includes the possible fabula.

In the case of non-fiction, the fabula is constrained by a claim to be the truth. While different interpretations of historical events, for example, can arise from weighting ambiguous evidence differently or from considering different perspectives, the fabula for a narrative history is chosen to best represent a particular interpretation[3]. In our model of situation augmentation (Figure 1), the fabula is determined by the information fusion system, possibly filtered for relevance and level of detail, but in any case separately from event sequencing. Removing the fabula from the plot solution space makes possible efficient algorithms which would not be suitable for the case of fictional narratives.

3 Problem Formulation

Assume the content to be conveyed is represented as a set of events:

$$C = \{E_k(t_k, \boldsymbol{e}_k, \boldsymbol{p}_k)\}, \quad 1 \leq k \leq n \tag{1}$$

where $\boldsymbol{e} = [e_1, e_2, \ldots]$ and $\boldsymbol{p} = [p_1, p_2, \ldots]$.

An event E is an n ary relation where the first term t is the time the event takes place. The term $\boldsymbol{e}_k$ is a list of those scenario elements that are affected by the event and the term $\boldsymbol{p}_k$ is a list of other parameters. For example: `course_change(0532, (Bismarck), (220°))` is an event at 0532 hrs affecting the element `Bismarck` with the parameter `220°` and `detected(0537, (Bismarck, Unidentified), (Port))` is an event at 0537 hrs affecting the elements `Bismarck` and `Unidentified` with the parameter `Port`.

The problem is to order the events E_i so as to best communicate the situation. The simplest solution would be to arrange the events in strict chronological order. Unfortunately, a complex evolving situation, involving many elements changing simultaneously, presented in strict chronological order can result in a disjointed narrative.

Anachrony, the telling of events out of chronological order [Genette, 1980], is very common in literature. Anachronies can be used to delay introducing details to the audience before they are needed, to motivate the audience by summarising significant events before detailing their causes, or even for deliberate ambiguity. Even though Genette devotes a great deal of attention to anachronies in the earlier parts of *Recherche du Temps Perdu* (the main subject of his analysis), he concedes that later 'the narrative in its major articulations, for the most part becomes regular and conforms to chronological order.' Sternberg [2004] argues that the significance of chronology has been overlooked by literary analysts both in terms of its prevalence and its literary significance. To the extent that anachronies are used in literature to produce suspense, plot twists, surprises and mystery, they are antithetical to the goal of enhancing situation awareness. In particular we wish to enhance the audience's ability to comprehend relationships between events (including temporal relationships), the intentions

[3] Histories which discuss alternate interpretations by evaluating the evidence would be considered *exposition* or *argument* modes of discourse rather than narratives, although they may contain narrative passages.

of characters and to 'project' the future course of events. For this application, a predictable, mimetic plot is desired. The challenge is how to maintain interest in such a plot.

Where different characters have different experiences, one common structure is to follow one character for a significant time before switching to follow another character for a significant time. Thus, in *The Lord of the Rings* [Tolkien, 1966], the party becomes split with Frodo and Sam heading to Mordor whereas Merry and Pippin are kidnapped and follow a very different path. This part of the novel has a macrostructure whereby alternate chapters deal with the adventures of the two groups. Following one party for a significant time encourages the reader to identify with the characters. Switching introduces a tension whereby the reader must delay the gratification of discovering the fate of the first party in favour of catching up with the fate of the second party. If the switching frequency is appropriate and happens at dramatic points then reader engagement is maintained or enhanced.

The goal of our current work is to use non-fictional narrative to enhance the audience's situation awareness rather than audience engagement per se. None the less, we expect engaged readers to be more attentive and attentive readers to achieve a higher level of awareness other things being equal[4].

We use the structure from *The Lord of the Rings* (discussed above) as inspiration for automated event sequencing. The structure may be characterised as having multiple story lines. Each story line is chronological but the story lines are asynchronous with respect to each other. Anachronies may occur at switches between the story lines. Story lines may intersect or merge at which point they become sychronised. We will refer to this structure as Multiple Intersecting Chronologies (MICs). The MIC structure results in a simple to comprehend narrative and is sufficiently constrained to permit tractable implementation. None the less it has sufficient flexibility to permit, for example, telling the background history of a new minor character to be deferred until the new character is needed. Additional flexibility can be achieved by using MICs as components of a more complex discourse structure. For example, a report containing a summary followed by elaboration can be treated as two narrations, each with a MIC structure, one with a low level of detail and the other with a high level of detail.

Considering only the ordering problem in constructing MICs, a balance must be struck between *continuity* and *temporal tension*. Continuity is the degree to which the narrative continues to follow the events experienced by one entity. Where events concerning disparate elements are occurring in parallel, continuity will necessarily result in events being narrated out of temporal order. Temporal tension refers to the degree to which events are presented out of time order. Since we expect that a fusion system will not be able to deduce all the causal relationships between events, especially in situations which are still evolving as the narrative is being constructed, it is important to preserve the ability of the audience to deduce causal relationships themselves. In effect, the machine constructs a partial story which the audience completes. If temporal tension becomes too great, the reader's situation awareness may be degraded due to achieving a false impression of what happened when.

In most narratives there will typically be some actors which are central (main characters) and some actors which are peripheral. In automatically constructing a narrative it may be

[4]There are caveats: for example including salacious descriptions of the characters private lives might increase audience engagement but distract from the pertinent aspects of the situation, or the audience might identify so much with particular characters that it creates a bias in the way they interprete events.

desirable to tell updates on the fate of central actors more often than less central actors so we introduce *centrality* as a third factor to be optimized. A narrative which focuses on central characters and defers telling the history of less central characters until they interact with more central characters, would exhibit high centrality and high temporal tension.

In Section 5, ways of estimating *continuity*, *temporal tension* and *centrality* are explored. In calculating these factors it is possible to take into account additional factors such as the nature of the event, the nature of the element and the operation under way. For example, an attack on a vital asset is more significant than a position update from a routine patrol.

There are clearly $n!$ ways of ordering the events in C, thus it is intractable to do an exhaustive search for the minimum cost ordering even for moderately large n (the Bismarck scenario used in this report contains 118 events).

4 Partial Temporal Ordering Algorithm

Total ordering based on the event times t_i can not be used as it may create disjoint narratives with insufficient continuity wheras an exhaustive search of all possible ways of ordering the events in C is intractable. However it is possible to reduce the complexity greatly by using the event time parameter to create a partially ordered set of events. The directed graph of events in Figure 2, has been drawn so that events occurring at a later time are to the right of events occurring at an earlier time. Edges indicate that the associated nodes have an element in common. Later events are assumed to have a *temporal dependency* on earlier events which have an element in common. Incorporating temporal dependency ensures the graph (Figure 2) is acyclic, which is greatly beneficial to efficient traversal of the graph. In the MIC structure, events involving a given element must be told in time order — the temporal dependency rule. To create a narrative with the MIC structure, we must traverse the graph without violating this rule.

Instead of operating directly on the set C, it is simpler to express the algorithm in terms of a related set

$$\hat{C} = \{R_{ik}\} \tag{2}$$

where R_{ik} are references to events. If an event has m elements, then m references are created

$$\left.\begin{array}{c} R_{1k} \\ R_{2k} \\ \vdots \\ R_{mk} \end{array}\right\} \rightarrow E_k(t_k, e_{1k}, e_{2k}, \ldots, p_{1k}, p_{2k}, \ldots). \tag{3}$$

The partial ordering is defined by $R_{ij} > R_{kl}$ if $e_{ij} = e_{kl}$ and $t_j > t_l$.

The references R can be arranged in queues, one for each element, sorted in time order (Figure 2).

The algorithm then consists of taking references off the queues. Each queue can be considered to represent the story line of an element — a chronology in the MIC structure. Optimisation is reduced to deciding which queue to choose next according to a suitable cost function

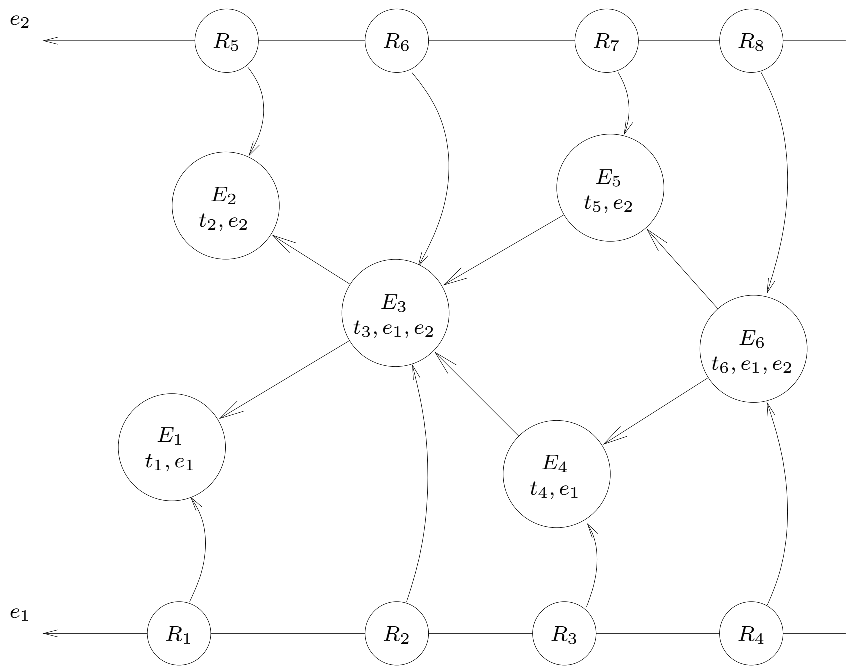

Figure 2: Mapping partially ordered events to queues of references. Solid edges indicate temporal dependencies between events E ($t_1 < t_2 < t_3 < t_4 < t_5 < t_6$). Broken arcs represent mapping from references R to events E. The numbering of the references is arbitrary.

(discussed further in Section 5). Additionally the algorithm must ensure that events with multiple elements (*nexus* events) are handled correctly. An element is said to be *ready* with respect to an event if its reference is at the head of a queue. A nexus is only ready to be told when *all* its elements are ready.

The algorithm is as follows:

1. Select the queue with the lowest value of the cost function.
2. If the reference at the head of the queue is to an event with one or more unready elements, then mark the queue as busy.
3. If the reference is to an event with no unready elements then tell the event. For all references to the event remove the references from their queues and mark the queues as not busy.
4. For all queues which are not busy, calculate the value of the cost function.
5. Select the queue with the lowest cost.
6. Repeat from step 2 while there are untold events.

The number of queues to consider at step 4 is likely to be significantly less than the number of events. Busy queues can be efficiently excluded from consideration. Queues consisting of a single nexus event are a degenerate case which can be eliminated without affecting the generated sequence of events. The elements represented by these queues are transitory: They take part in the story lines of other elements but have no chronology of their own. Queues which become empty can also be eliminated from further consideration. In the Bismarck scenario, there are 46 elements and 118 events. Of the 46 elements, 19 are degenerate leaving at most 27 queues to consider. It is possible to conceive of narratives with very large numbers of actors. For example there were over 2000 crew on the Bismarck alone, however these actors are treated as a collective ('the crew'). They can only be given individual story lines if they are differentiated by the events that affect them. The number of distinguishable elements is roughly bounded[5] by the number of events and in practice we expect it to be significantly less. The algorithmic complexity is reduced further if the cost function is dependent only on the events referred to at the head of the queue.

This algorithm is an efficient implementation of a constrained optimization problem. Organizing the event references into queues enforces the temporal dependency constraint. The constraint provides efficiency by reducing the dimensionality of the optimization space.

5 Cost Function

We calculate the cost of selecting the next queue from measures of the *Temporal Tension*, *Centrality* and *Continuity* concepts (introduced earlier in Section 3). In the case of temporal tension and continuity, we have devised alternative measures. We treat the choice of measures and their relative weighting as questions to be resolved empirically (Sections 6 and 7).

We need to rely on parameters which can be machine generated. Since we have chosen to partition the overall system into a fusion system and an IMMP interface, it is the responsibility of the fusion system to generate these parameters and the role of the IMMP system to best use these parameters to generate a presentation.

The time scale in which events occur (tempo) can vary greatly even within one narrative. Thus a temporal tension of days may not be significant in a slowly moving narrative whereas a temporal tension of seconds may be significant if it is important to know, for example, who shot first in a battle. One way to allow for this elasticity of time, is to measure temporal tension by the number of events which would have to be told to bring the elements up to date with respect to the selected element. However, all events do not contribute equally to the temporal tension. Especially with machine reported events from radar, transponder or other surveillance technology, there could be a large number of events which are simply position updates. We have devised three alternative measures of temporal tension.

The *temporal tension (absolute)* is the time difference between the head event of an element queue and the earliest unhandled event[6]. The cost of choosing the i^{th} element due to the magnitude of unhandled time is given by $\tau_i = t_i - t_{\text{earliest}}$ where t_i is the time of the first event of the i^{th} element queue and t_{earliest} is the time of the earliest unhandled event in all elements. This method of estimating temporal tension is sensitive to variation in tempo.

The *temporal tension (significance)* is the sum of the significances of all events which occur earlier in time than the head of a given queue. Significance is a feature discussed earlier (Section 1) in the context of content selection. It is something that can be estimated by

[5] Strictly it is possible to have $2^n - 1$ distinguishable chronologies through n events however we consider this worst case highly unlikely.

[6] In literary analysis terms, this is the *span* of the anachrony [Bal, 2009].

the fusion system using knowledge of platform capabilities, cost, vital asset classification, adversary intent and current plans although in the current implementation significance has been assigned based on subjective evaluation by the researchers. Important events such as the sinking of the Hood or the rudders becoming jammed on the Bismarck are assigned high significance, while events such as a maintaining a course are assigned low significance.

The significance tension for the i^{th} element is given by

$$\hat{s}_i = \sum_{n=0}^{j} S_n \qquad \text{where} \quad j = \arg\max_j \left(t_j | E_j \in C; t_j < t_i\right) \tag{4}$$

and S_n is the significance value of the n^{th} event. This method of estimating temporal tension is independent of variations in tempo.

These two methods of determining temporal tension can be combined by taking their product. This *combined* method is simple, and provides a compromise between tempo sensitive and tempo independent means of estimating temporal tension.

Narratives typically revolve around *main characters*. While the algorithm (Section 4) will naturally result in elements involved in many events being visited more often, it may in addition be desirable to bias the algorithm towards choosing central elements more often.

Consider a network in which each element in the narrative is a node and each event is an edge, as in Figure 3.

The principle of centrality is a measure commonly used to describe how important or powerful a node is in a network, [Monge and Contractor, 2003]. There are many different types of centrality, the simplest is degree centrality, which is calculated by counting the number of nodes directly connected to a node [Monge and Contractor, 2003]. So for example in Figure 3 the degree centrality for all of the elements is 3 as they are all connected to each other. The minimum degree value that an element which is in at least one event can have is 1 as it is considered to be connected to at least itself. In the Bismarck scenario, the most central elements by this measure are[7] the Bismark (1·0), the British Admiralty (0·46) and Admiral Tovey (0·38).

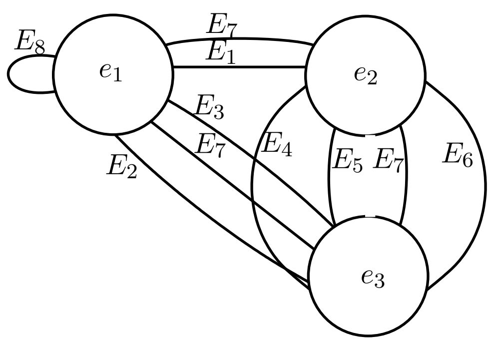

Figure 3: Element relatedness as a network.

[7]The values in parenthesis are normalised on the interval $[0, 1]$.

When two elements are closely related (for example members of a party or antagonists exchanging fire) telling the story line of one element is likely to also advance the story line of the other. Thus switching to a closely related element may result in a smaller break in continuity than switching to a unrelated element.

The relatedness of each element to each other element can be extracted from the network of elements, based on the number of events in which two elements both occur. A given element e_i belonging to events $[E_{i1}, E_{i2}, \ldots, E_{in}]$ is related to an other element e_r by the measure w_{ij} given by:

$$w_{ij} = \frac{1+l}{1+n} \tag{5}$$

where

$$l = \sum_{\substack{1 \le h \le n \\ 1 \le k \le n}} r_{hk}$$

and

$$r_{hk} = \begin{cases} 1; & E_{ih} = E_{jk} \\ 0; & E_{ih} \neq E_{jk} \end{cases}$$

The +1 term in the numerator and denominator of Equation 5 was introduced to avoid the possiblilty of a zero denominator[8].

In network analysis terms l is the number of edges connecting the nodes e_i and e_j and n is the total number of edges connected to node e_i. For example, in Figure 3, $w_{1,3} = \frac{1+3}{1+6} = 0{\cdot}57$ as there are 3 edges connecting e_1 and e_3, and e_1 has a total of 6 edges connected to it. It should be noted that every element has maximum relatedness to itself as it shares all of its events with itself. Relatedness is a whole of narrative property. Thus two elements can be strongly related even though at a particular point in the narrative they are not interacting.

Two different methods of calculating *continuity* have been implemented. In the binary method, the continuity is 1 if the event shares any elements with the last event or 0 if it doesn't. The binary method allows switching between queues at a nexus with no penalty. Thus, for example: 'The man picked up a shoe; He threw it at the cat; The cat ran away;' has no break in continuity under the binary method even though the first event and the last event have no elements in common.

The other method uses the relatedness (Equation 5) between the element associated with the last queue and the element associated with the candidate new queue as the measure of continuity. The relatedness method is never used alone. When enabled, it is only invoked for events which are discontinuous under the binary method. In the Bismarck scenario (for example), the Norfolk is most closely related to the Bismark (0·63), the Suffolk (0·5) and the Prince of Wales (0·5). The Norfolk and Suffolk are operating in concert to detect and shadow the Bismarck. Choosing the relatedness method favours switching focus from the Norfolk to the Bismarck or from the Norfolk to the Suffolk even when there is no nexus affecting both ships.

[8] With the current implementation n is greater than zero for all elements but with earlier implementations, elements with no events where possible.

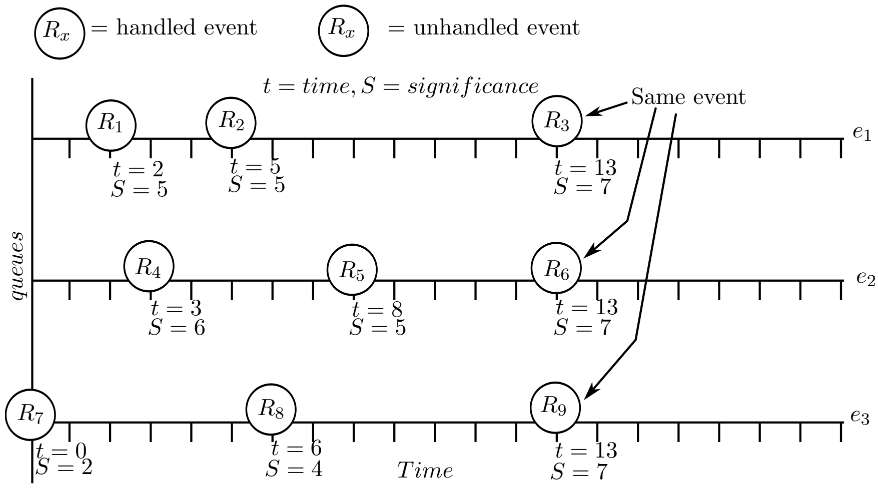

Figure 4: An example scenario in which a narrative is spread across 3 elements e_1, e_2 and e_3. R_1 refers to an event involving e_1, which has just been handled. R_{2-9} are events yet to be handled. The queues of references are labeled e_{1-3}.

Figures 3 and 4 illustrate a scenario in which the event that is referenced by R_1 of element e_1 has been handled. The possibilities at this point in processing are to continue with queue e_1 in which case the next element is referenced by R_2 or to switch to queues e_2 or e_3 in which case the events referenced by R_4 or R_7 respectively are next to be processed. Table 1 shows the calculated values of various features for each of choices. Figure 4 was used to gain the *temporal tension* costs and Figure 3 was used to obtain the relatedness and degree values. In order to obtain a total cost the values are normalised, weighted and summed. Since we postulate that *temporal tension* is a cost to be minimised whereas *continuity* and *centrality* are both factors to be maximised (representing a reward), the total cost G is given by:

$$G = \alpha\hat{T} - \beta\hat{N} - \gamma\hat{D} \tag{6}$$

where $\hat{T}, \hat{N}$ and $\hat{D}$ are the normalised results of the tension, continuity and centrality costs and α, β and γ are the weights.

Table 1: Values for each element in Figure 4

Feature Calculated	e_1	e_2	e_3
Temporal tension (absolute)	$5 - 0 = 5$	$3 - 0 = 3$	$0 - 0 = 0$
Normalised	$\frac{5}{5} = 1$	$\frac{3}{5} = 0{\cdot}6$	$\frac{0}{5} = 0$
Temporal tension (significance)	$6 + 2 = 8$	2	0
Normalised	$\frac{8}{8} = 1$	$\frac{2}{8} = 0{\cdot}25$	$\frac{0}{8} = 0$
Continuity (binary)	1	0	0
Continuity (relatedness)	$w_{1,1} = 1$	$w_{1,2} = 0{\cdot}43$	$w_{1,3} = 0{\cdot}57$
Degree Centrality	3	3	3
Normalized	$\frac{3}{3} = 1$	$\frac{3}{3} = 1$	$\frac{3}{3} = 1$

Table 2: Key definition for referring to algorithm methods and weights.

Symbol[a]	Description	Values
x	Temporal tension (absolute) Temporal tension (significance) Temporal tension (combined)	A S C
y	Continuity (binary) Continuity (relatedness) Continuity (equivalent results)	B R E
α	Temporal tension weight	$0 \leq \alpha$
β	Continuity weight	$0 \leq \beta$
γ	Centrality weight	$0 \leq \gamma$
Example	Temporal Tension (absolute) Continuity (relatedness) Temporal tension weight: 0·1 Continuity weight: 0·3 Centrality weight: 0·5	A-R,0·1,0·3,0·5

[a] Key is of the form x-y,α,β,γ. Permissible values are as tabulated.

It can be seen that by using different combinations of methods and weights the system can be tuned to produce a variety of narratives within the limits of the MIC structure (enforced by the temporal dependency rule). High α (temporal tension weight) with low β and γ produces a narrative which tends towards a strict chronology. High β (continuity weight) and low α and γ tends to produce a narrative which follows the action from cause to effect. High γ (centrality weight) with low α and β tends to produce a narrative which follows the central character and defers telling the background of secondary characters until they interact with a more central character.

In order to concisely refer to the methods and weights we introduce a key of the form x-y,α,β,γ where x refers to the method for determining temporal tension, y refers to the method for determining the continuity and α, β and γ are the weights in Equation 6. Permissible values for these parameters are documented in Table 2.

6 Implementation and Testing

The implementation of this algorithm was written in C++. The program takes input from a text file (see Appendix A) that contain the events one event per line.

As the input file is processed, elements and events are placed in vectors. Elements and events are referred to by their index in these vectors. For each new element encountered, an empty queue is created. When the event vector has been fully populated, it is traversed in time order adding event references to the queues corresponding to each element.

With the events and references set up appropriately, the initial event is chosen, based on just *temporal tension* and *centrality*, as *continuity* for the initial event is meaningless. The program then progresses through the events as stated by the algorithm in Section 4. Rather than minimising the cost (Section 5), the program maximises the reward, $P = -G$. It is convenient for all the terms of P to have the same sign, which can be achieved by setting $\tau = -T$. Thus the goal becomes to maximise

$$P = \alpha\hat{\tau} + \beta\hat{N} + \gamma\hat{D}$$

This ensures that the total reward is always positive. The concept of a 'busy' element is realised by setting the reward value of that element to -1. As the algorithm runs, statistics are gathered so that the performance of the program can be evaluated.

The algorithm was tested on a set of events obtained by de-constructing a narrative of the Bismarck scenario. The events included a text snippet (from the source material) describing the event in the parameter list $\boldsymbol{p}$. Thus, when the events have been suitably ordered, the text snippets can be concatenated to produce an understandable narrative, albeit lacking the naturalness which implementing the finishing steps would bring.

The generated sequence of events were compared to the sequences of events in a corpus of three human authored texts by way of the common natural language generation (NLG) metric edit or Levenshtein distance. The Levenshtein distance was chosen over other natural language metrics such as BLEU (Bilingual Evaluation Understudy) or NIST [9] as they are designed to account for variations in linguistic realization seen in machine translation. Since the system will be run with a fixed set of events and only the order of these are being considered, the Levenshtein distance was considered more appropriate. The human authored and machine generated narratives were also compared on the values of the temporal tension and continuity metrics. Finally the machine generated text was manually examined for cases where the narrative appeared disjoint or where comprehension may be compromised. These cases are presented and discussed in Section 7.2.

6.1 Source and human authored texts

The narrative chosen was the Operation Rheinübung and sinking of the battleship Bismarck. Operation Rheinübung was an attempt by the German battleship Bismarck and heavy cruiser Prinz Eugen to breakout into the Atlantic and attack supply lines during World War II. During the operation the British battlecruiser HMS Hood was sunk, this prompted a large retaliation and hunt for the Bismarck which ended in her sinking. This was a large complex operation involving nearly 100 ships [Wikipedia, 2014b] (although the subset of events we have used involved only 46 elements). As an historical naval scenario there is a wealth of information freely available including ships logs with accurate time and position data. Although large naval battles are not a feature of modern warfare this scenario is still militarily relevant with respect to generating situation narratives because it is dominated by events and contains multiple actors with antagonistic goals.

To validate the results objectively the output was compared to a corpus of human authored texts. The corpus consisted of:

[9] NIST refers to a machine translation evaluation metric developed by the US National Institute of Standards and Technology, however it does not appear to have the status of a formal standard. It is related to BLEU.

- Full set - generated from Asmussen [2013b] and Asmussen [2013a]
- Subset KBismarck - generated from Rico [2013]
- Subset Wikipedia - generated from Wikipedia [2014c], Wikipedia [2014b] and Wikipedia [2014a]
- Subset NavWeaps - generated from DiGiulian [1994].

For each of the human authored texts it was necessary to locate the descriptions of previously defined events within that text. This was done so that an ordered list of events for that text could be derived. With the ordered list of events various statistics were obtained, such as the amount of temporal tension present in the human authored text. Different authors select different subsets of the events in order to relate the narrative in their own way. Since this work was concerned with event ordering and not event selection, for each comparison, only the subset of events found in the particular human authored text were used. The output of the system was compared to the human authored event subset via the use of the common natural language generation metric, edit (or Levenshtein) distance. It was also noticed that in several cases the human authored text appeared to break the temporal dependency rule, however this may be due to occasional ambiguity as to which event the description referred to. Each of these texts have features about them worth noting.

KBismarck: This document is the most thorough of the documents and contains the most details. It also seems to be told predominantly from the point of view of the Bismarck, as its name would suggest, which could affect the order in which it presented events and which elements it considered to be main characters. Despite this it seemed to be the most engaging of the documents. It also contains substantial amounts of discussion on the decisions made by the officers on each side.

Wikipedia: The Wikipedia subset was actually synthesised from three smaller documents, one on the battle of the Denmark strait, one on Operation Rheinübung and one on the last battle of the battleship Bismarck. It was considered that these documents be read in this order but each was written as a stand alone document. Events repeated in documents were ignored. Despite this it provides a quite engaging read and still covers much of the detail. Similar to the KBismarck it touches on a lot of the decisions made by the officers, but unlike the KBismarck it is less focused on the Bismarck and switches perspectives more often.

NavWeaps: This document appears to be a more technical summary of the encounter rather then a listing of events. It contains many diagrams and tables about the ship specifications and also contains accounts of an expedition to the wreck. It contains a large amount of detail for the battles and engagements but does not contain as many events as the other two. This document achieved the least engagement perhaps due to its focus on ‘dry’ technical details.

6.2 Parameters

To test this implementation, the algorithm was run multiple times using different methods and weights (Equation 6) to calculate the cost.

There are three temporal tension methods (absolute, significance and combined) and two continuity methods (binary and relatedness). The values of the weights α, β and γ can also be changed from test to test. The weights were assigned the values $0{\cdot}1, 0{\cdot}5, 1, 7, 15$. Each of these combinations of parameters were tested with three different content subsets associated with the different human authored texts.

We wish to determine the optimal combination of methods and weights and testing all the combinations for all the sources results in a total of $3 \times 2 \times 5^3 \times 3 = 2250$ trials.

7 Results and Discussion

7.1 Statistics

Evaluation of a configurations success was based on how similar the statistics it produced were to the statistics calculated for the human authored texts (Table 3).

From Table 4 it can be seen that several of the configurations produce results with similar statistics to the KBismarck human authored text, which was subjectively the best of the three. There is no configuration which minimises all three statistics (edit distance, and relative difference by both absolute and significance methods). However low values of all three statistics were achieved with the absolute method (row 3), the significance method (row 5) and the combined method (row 8) for calculating the tension cost and both the binary and relatedness methods (rows 3 and 5) for calculating the continuity cost.

The result achieved by the C-R,7,0·5,1 method (row 8) produces statistics very similar to those in the KBismarck text and while it has an Edit distance of 29, slightly higher than most of the others, it is still a lot lower than the worst case of 83 (not included in Table 4). This provides support for the inclusion of the combined method (absolute and significance) for calculating temporal tension and the relatedness method for calculating continuity.

It is also interesting to note that the method which produced the lowest edit distance achieved was by a run (row 7) that also produced strict chronological output (zero total temporal tension). Lower total temporal tension usually results in a lower edit distance, which is to be expected since the human authored input is predominately in chronological order with relatively few exceptions. As the weights are adjusted to achieve a better match between the temporal tension in the human authored and the machine generated discourse, there is more variation permitted in the machine generated discourse and so the edit distance increases.

Table 3: Statistics derived from the human authored texts.

Text Source	Temporal Tension (absolute)	Temporal Tension (significance)
KBismarck	1873·1	157
Wikipedia	10023·4	239
NavWeaps	1014·9	159

Table 4: Comparison of machine authored text with KBismarck human authored text by edit distance and temporal tension statistics — selected results.

			Temporal Tension			
		Edit	Absolute		Significance	
	Method[a]	distance	Total	RD[b]	Total	RD[c]
1.	A-E,15, 0·1, 0·1	24	41·1	97·8%	51	67·5%
2.	A-E,7, 0·1, 1	28	1946·9	3·9%	126	19·7%
3.	A-E,15, 15, 0·1	28	1683·1	10·1%	156	0·6%
4.	S-E,15, 0·1, 0·1	20	180	90·4%	3	98·1%
5.	S-E,7, 0·5, 0·1	26	1888·9	0·8%	58	63·1%
6.	S-B,1, 0·1, 0·1	28	6791·1	262·6%	153	2·5%
7.	C-E,15, 0·1, 0·1	19	0	100·0%	0	100·0%
8.	C-R,7, 0·5, 1	29	1866·1	0·4%	136	13·4%
9.	C-E,7, 1, 1	30	2464·1	31·6%	156	0·6%

[a] Method key: refer to Table 2.
[b] Relative difference to KBismarck result (Table 3) of 1873·1.
[c] Relative difference to KBIsmark result (Table 3) of 157.

These results demonstrate that the parameters can be tuned to produce similar statistics to human authored text but narratives with similar statistics can still differ significantly as indicated by the associated edit distances. However, a different event order does not necessarily mean a worse narrative: it is possible that the machine generated event order is actually better than that of the human authored narrative; or that the machine generated event order is different but of the same quality, just as there are differences in event order between human authors (perhaps to achieve a desired affect or just as personal style preference).

7.2 Vignettes

Tables 5–10 shows the output of the system given various inputs. Table 5 shows a portion of output that experiences positive results due to event ordering. By presenting the second Prinz Eugen event out of order (before the first Hood event rather than after) it keeps the events relating to what the Prinz Eugen is doing together, then presents the events relating to what was happening on the Hood together. This is an example, we consider, where high continuity weight results in improved text quality.

Table 6 shows the same section as Table 5 but with different settings. This section is still an example of a positive result, but in this example the event regarding the Prinz Eugen changing targets is deferred. This is because the systems higher centrality weighting results in a tendency to group events containing the Bismarck together. At the start, events regarding what happened to the Hood are presented and then events based on the Prince of Wales are grouped together.

This is interesting because it contains the same events as Table 5 but by ordering the events

Table 5: Section of output produced by: S-E,15,1,0·5; A-E,7,1,0·1 (Table 2) and others.

Time	Element	Event description
694	Prinz Eugen	The Prinz Eugen hits the Hood near the hangar by the after mast with its 4th salvo
697	Prinz Eugen	Lutjens commands the Prinz Eugen to change its target to the Prince of Wales
694·1	Hood	A fire starts on the Hood near the port side ammunition lift
698	Hood	The Hood changed course by 20 degrees to bring the after turrets into action
699	Bismarck	After the Bismarck's 5th salvo the Hood blows up
700	Bismarck	The Bismarck changes target to the Prince of Wales
700·1	Prince of Wales	The Prince of Wales turns hard around the Hood wreckage

differently actually leads to a slightly different interpretation. The narrative is, in the first case, about the Prinz Eugen then the Hood and in the second case, about the Hood and then the Prince of Wales. While this reads nicely and is engaging there is a bit of contention amongst historians about which ship fired the shot that sunk the Hood, the main evidence for one opinion is that the Prinz Eugen had changed targets before the Hood sunk, and obviously when the items are presented in this order it could affect the audiences understanding of the situation. In cases like this if the exact order of events is not maintained, situation awareness would be degraded.

Table 6: Section of output produced by: C-E,7,0·5,1; S-E,7,1,1 (Table 2) and others.

Time	Element	Event description
694	Prinz Eugen	The Prinz Eugen hits the Hood near the hangar by the after mast with its 4th salvo
694·1	Hood	A fire starts on the Hood near the port side ammunition lift
698	Hood	The Hood changed course by 20 degrees to bring the after turrets into action
699	Bismarck	After the Bismarck's 5th salvo the Hood blows up
700	Bismarck	The Bismarck changes target to the Prince of Wales
700·1	Prince of Wales	The Prince of Wales turns hard around the Hood wreckage
697	Prinz Eugen	Lutjens commands the Prinz Eugen to change its target to the Prince of Wales

Table 7 shows a section of output of events which occur right after the events in Tables 5 and 6. In this case the system tries to continue with events concerning the Bismarck for too long.

Table 7: Section of output produced by: C-E,7,0·5,1; S-E,7,1,1, S-E,15,1,0·5; A-E,7,1,0·1 (Table 2) and others.

Time	Element	Event description
700·1	Prince of Wales	The Prince of Wales turns hard around the Hood wreckage
701	Bismarck	The Prince of Wales remains under concentrated fire from both the Bismarck and the Prinz Eugen
708	Bismarck	The Bismarck was hit a total of 3 times causing a reduction in speed and a large oil spill
1059	Bismarck	At noon the German ships set a new course of 180 degrees
702	Prince of Wales	The Prince of Wales took a direct hit in the bridge
702	Prince of Wales	Rear Admiral Walker took over further measures as Senior Commander
702·1	Prince of Wales	The Prince of Wales left the battlefield smoking heavily
704	Prince of Wales	The Prince of Wales was hit a total of 7 times

This causes the event at time of 1059 to be presented out of time order. It isn't the magnitude of this out of ordering that is bad, but the fact that it is out of place and doesn't make much sense in relation to the events around it. This shows that while one set of weight values can produce good results for a specific group of events and then bad results for another, suggesting that weights assigned adaptively based on the local pace of the narrative could help to avoid this issue.

Table 8 shows another undesirable section of output that was commonly produced. Once again continuity is maintained for too long resulting in the system attempting to follow the idea for an air strike but in doing so causes the sequence of events relating to how the Bismarck provides a distraction (time 1433) to enable the Prinz Eugen to disengage and move into the Atlantic (also time 1433) to be split up and lose continuity. This is because the event of the Prinz Eugen 'leaving formation' does not capture the fact that the formation also involves the Bismarck and the event of the Bismarck's manoeuvre does not capture the intention to provide a diversion for the Prinz Eugen.

Table 9 shows how the same events are dealt with by a different configuration. By presenting the event at time 1459 out of time order it shows that the Bismarck performs a distraction, and that this enables the Prinz Eugen to escape more clearly than if it wasn't presented out of order. It also doesn't move away from the lead up to and implementation of the air strike too much.

Table 10 contains an example which shows the affect of an error in preparing the data. The event at time 1954 only contains the Bismarck element, and when the event is naively examined it can be read as the Bismarck makes a complete circle, in reality however the event relates to the Bismarck escaping from her pursuers which would mean that it should

Table 8: Section of output produced by: C-E,7,1,1; A-E,1,0·1,0·1 (Table 2) and others.

Time	Element	Event description
1134	British Admiralty	In response to the sinking of the Hood, the British Admiralty marshals all its resources. It sends cruisers Manchester, Birmingham and Aresthusa to patrol the Iceland Faroes gap
1134	British Admiralty	Rodney was diverted to join Tovey's force
1248	Bismarck	Tovey orders Victorious, Galatea Aurora Kenya and Hermione to close range to the Bismarck and, prepare to launch an air strike
1433	Bismarck	Later the Bismarck turns 180 degrees to starboard toward her British pursuers
1459	Bismarck	The Bismarck fires upon the Suffolk and Prince of Wales killing 3 sailors
1749	Bismarck	That night the carrier Victorious launched an air strike on the Bismarck
1759	Bismarck	During the attack the Bismarck was hit by a torpedo on the starboard side without importance
1134	British Admiralty	Ramillies was ordered to take a position west of the German battleship
1134	British Admiralty	Revenge was sent on a course to intercept
1433	Prinz Eugen	The Prinz Eugen left formation and headed into the Atlantic

Table 9: Section of output produced by: A-R,15,1,0·5 (Table 2) and others.

Time	Element	Event description
1248	Bismarck	Tovey orders Victorious, Galatea Aurora Kenya and Hermione to close range to the Bismarck and, prepare to launch an air strike
1433	Bismarck	Later the Bismarck turns 180 degrees to starboard toward her British pursuers
1459	Bismarck	The Bismarck fires upon the Suffolk and Prince of Wales killing 3 sailors
1433	Prinz Eugen	The Prinz Eugen left formation and headed into the Atlantic
1749	Bismarck	That night the carrier Victorious launched an air strike on the Bismarck
1759	Bismarck	During the attack the Bismarck was hit by a torpedo on the starboard side without importance

Table 10: Section of output that is produced by most configurations

Time	Element	Event description
1954	Bismarck	While the ships shadowing it were zig zagging the Bismarck makes a complete circle to lose them
1955	Bismarck	The Bismarck turns toward starboard with a new course of 130 degrees
1895	Prince of Wales	Due to now being in an area with high risk of U-boat attack the Norfolk Suffolk and the Prince of Wales start to zig zag
1955	Bismarck	The Suffolk loses contact with the Bismarck, contact lost for 31 and a half hours

also contain the elements Suffolk, Norfolk and Prince of Wales. This is noteworthy because it is the kind of error the fusion system could easily make, picking up on the obvious physical event and missing its larger context, and it results in output which could well mislead the audience.

7.3 Discussion

There are two working hypotheses proposed in this report:

1. exploiting partial ordering of events leads to efficient discourse planning
2. minimising a weighted sum of *continuity*, *temporal tension* and *centrality* results in an engaging narrative.

However there are also a number of assumptions made in the design of the algorithm. The degree to which this algorithm is useful is largely dependent on the degree to which these assumptions hold:

1. It has been assumed (Section 1) that the fusion system and the IMMP system can function independently. The concerns of the fusion system and the IMMP system can be separated with the fusion system having deep knowledge of elements and events but shallow knowledge of the audience, media and story telling; Whereas the IMMP system would have deep knowledge of the audience, media and the art of story telling but only shallow knowledge of elements and events.
2. It has been assumed that the content to be conveyed is contained in a set of events as defined in Section 3. While there is considerable flexibility afforded by the event parameters p_{ik}, which can contain arbitrarily complex information, events are constrained in our representation to occur in an instant at time t_k. However, real events may have a significant duration, or even if the event occurs effectively instantaneously, it may only be certain that an event occurred at some point in a time *interval*.
3. It has been assumed (Section 4), that events are temporally dependent on earlier events referring to the same elements and that all relevant causal relationships are exposed

through these temporal dependencies. While assuming an irrelevant temporal dependency is relatively harmless, potentially resulting in a correct but less engaging result, missing dependencies (not identified by the fusion system) can result in incorrect misleading output. Further is is possible that the fusion system may be able to deduce that one event must have occurred after another without knowing the precise time of either, suggesting that a way of conveying temporal dependency independently of absolute times t_k is needed.

4. It has been assumed that plot generation can be performed with only a shallow understanding of the content to be presented. The cost function we have developed (Section 5) relies only on the time of the events, the elements related by the event and the significance (assumed to be assigned by the fusion system) of the events. The discourse planning makes no use of knowledge of the nature of events, their impact, their causal relationship or the nature of elements, their capabilities and ontological relationships etc.

5. It has been assumed (Section 1) that determining the order in which events should be told can be performed independently of other functions of the IMMP including other aspects of discourse planning. In the Bismarck scenario, the human authored text includes a considerable amount of supplementary information beyond the actual events. For example, contextual information includes strategic, geo-political, geo-spatial, astronomical and meteorological aspects as well as platform capabilities which degrade over time due to fuel consumption and battle damage. Switches of story line from one element to another could provide a suitable cue for the introduction of supplementary information. Thus context for the next segment can be introduced immediately after a switch in story line and summarization can be added immediately before a switch in story line.

The testing of the algorithm has yielded significant insight into its characteristics and limitations, however it can not be considered a rigorous evaluation. The output was evaluated by comparing statistics with human authored text, but this leaves unanswered the question of whether these are the relevant statistics. It would be more valid to assess audience attention, recall and comprehension or their ability to perform a relevant task although such methods are expensive and often impractical [Reiter and Belz, 2009]. Only one scenario (albeit a large one) has been used leaving open the question of whether the algorithm is generally applicable and whether the parameters α, β and γ need to be tuned for each scenario. With these limitations, testing has provided support for our hypotheses.

With this size of scenario (118 events) computation time was insignificant. The only constraint imposed by the partial ordering aspect of this algorithm is to enforce the temporal dependency rule. Since the human authored text always followed the temporal dependency rule, hypothesis 1 is supported.

Inspection of the vignettes (Tables 5 – 10) demonstrates that varying the weights for *continuity*, *temporal tension* and *centrality* provides sufficient scope to significantly affect the discourse. Discourse similar to the human authored output can be achieved with suitable weights lending support to hypothesis 2. The best results were achieved with non-zero weights for all three features giving further support for this hypothesis.

In assessing whether *continuity*, *temporal tension* and *centrality* are the appropriate features to use in discourse planning, is is difficult to separate the concepts from the implementation. The methods we have developed for estimating these measures are constrained by assumptions 2 and 4. Better means of estimating these measures, which may require abandoning these assumptions, could lead to significant improvements.

Relatedness (used to estimate *continuity*) depends only on the number of events shared by two elements, however two closely related elements (for example members of a fleet) may never be explicitly mentioned in the same event. To an extent this could be an artefact of the information sources[10]. An alternative would be to use an ontology to reason about relatedness, for example, reasoning that the activities of a fleet affect all the members of the fleet. This reasoning might be best done by the fusion system (maintaining assumption 1) but a richer representation of the content would be required to convey this information to the IMMP system.

Estimating *temporal tension* by the significance method has the advantage that it is scale independent. However the best performance was achieved with the combined method which is still dependent on time-scale indicating that there is scope to find a better measure of temporal tension.

Centrality calculated by degree successfully identifies the Bismarck as the central character in the test scenario. Whether this method is applicable to other scenarios has not been tested.

Two vignettes (Section 7.2) show the potential for event ordering to affect the interpretation of the situation by the audience. In the first case (Tables 5 and 6), whether the Prinz Eugen or the Bismarck fired the shot that sunk the Hood is not known. These are actually alternate scenarios (albeit with no difference in outcome). The source data does not explicitly represent either scenario and is in this respect incomplete. The second case (Table 10) regarding the nature of the Bismarck's deception manoeuvre was the result of a naive interpretation of the source text, but the effect was that the set of events fed to the machine was over-simplified and thus incomplete.

No fusion system fed with incomplete and uncertain data can be expected to produce accurate assessments of events. The problem is that presenting the events out of order, as permitted by the information available, may mislead the audience and actually hinder them in forming their own more correct awareness.

This is a special case of a more general problem: *Presenting incomplete and uncertain information, as if it is complete and certain is misleading.* This applies to human analysts as well as to machine fusion systems, but it does highlight the need for situation awareness aids to represent and present incompleteness and uncertainty.

[10] If all the intra fleet messages and commands were included, the relatedness of the members of the fleet would be exposed.

8 Conclusion

An efficient plot generation algorithm has been developed for situation narratives. The problem is to order a given set of events so as to maximise audience engagement while not misleading the audience about the temporal relationship between events.

The algorithm can be viewed as a constrained optimisation problem where a weighted sum of measures of *continuity*, *temporal tension* and *centrality* gives a cost to be minimised subject to a temporal dependency rule. The temporal dependency rule permits events to be told out of order but imposes a partial order where events concerning a given element must be presented in time order.

The algorithm has been tested using the events leading up to the sinking of the World War II battleship Bismarck as a scenario. With appropriate weights, machine generated discourse was comparable with human authored narratives relating the same events. A more rigorous assessment would include testing with a variety of scenarios and assessing audience attention, recall and comprehension.

The algorithm's computation was insignificant for the given scenario and the results support the use of *continuity*, *temporal tension* and *centrality* measures in discourse planning. There is scope to further improve the algorithm by devising better measures. Currently the best performing method of measuring temporal tension is sensitive to the tempo of operation. If the algorithm is to be generally applicable to a wide variety of scenarios without tuning, then a measure of temporal tension which gives high performance and is independent of tempo needs to be devised.

When the input event set did not properly represent the temporal dependencies, the algorithm was observed to produce misleading output. Human analysts are susceptible to the same problem, but it may be that human analysis have access to other information (including implicit real world knowledge) which mitigates the problem.

This work has assumed that determining the order in which events should be related is independent of determining how those events should be told. This assumption is largely untested. The major area of future work is to integrate this algorithm into a complete system. Even if the independence assumption does not hold, we postulate that measures of *continuity*, *temporal tension* and *centrality* will remain valuable criteria for optimization and that using partial ordering based on temporal dependency to constrain the solution space will lead to more efficient algorithms than would otherwise be the case.

9 Acknowledgements

The authors would like to acknowledge the comments and suggestions from Steven Wark and Adam Saulwick which have helped improve this report.

10 References

Asmussen, J. (2013a) *Operational History*. Bismarck Website. manual. Accessed 2013.12.09. Bismarck.com.
URL: http://www.bismarck-class.dk/bismarck/bisoperationhist.html.

— (2013b) *The History of the Bismarck*. Bismarck Website. manual. Accessed 2013.12.09. Bismarck.com. URL:
http://www.bismarck-class.dk/bismarck/history/bisdenmarkstraitbattle.html.

Bal, M. (2009) *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press. ISBN: 9780802096319.
URL: http://books.google.com.au/books?id=jPj4Bq0H4JoC.

Bui, V., Abbbass, H. and Bender, A (2010) 'Evolving stories: Grammar evolution for automatic plot generation'. In: *Evolutionary Computation (CEC), 2010 IEEE Congress on*, 1–8.
DOI: 10.1109/CEC.2010.5585934.

Bui, V., Bender, A and Abbass, H. (2012) 'An expressive GL-2 grammar for representing story-like scenarios'. In: *Evolutionary Computation (CEC), 2012 IEEE Congress on*, 1–8.
DOI: 10.1109/CEC.2012.6252888.

DiGiulian, T. (1994) *Bismarck's Final Battle*. Website. manual. Accessed 2014.01.09. International Naval Reseach Organization and Nav Weaps.com.
URL: http://www.navweaps.com/index_inro/INRO_Bismarck_p1.htm.

Endsley, M. R. (1988) 'Design and Evaluation for Situation Awareness Enhancement'. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **32** (2), 97–101. DOI: 10.1177/154193128803200221.
eprint: http://pro.sagepub.com/content/32/2/97.full.pdf+html.
URL: http://pro.sagepub.com/content/32/2/97.abstract.

Genette, G. (1980) *Narrative discourse : an essay in method*. Cornell University Press. Ithaca, N.Y. ISBN: 0-8014-9259-9.

Genung, J. (1900) *The Working Principles of Rhetoric Examined in Their Literary Relations and Illustrated with Examples*. Ginn.
URL: http://books.google.com.au/books?id=MkhKAAAAIAAJ.

Gervas, P. et al. (2005) 'Story plot generation based on CBR.' In: *Knowledge-Based Systems* **18** (4), 235 –242. URL: http://search.ebscohost.com/login.aspx?direct=true&AuthType=athens&db=inh&AN=8742147&site=eds-live.

Kybartas, B. and Verbrugge, C. (2014) 'Analysis of ReGEN as a Graph-Rewriting System for Quest Generation.' In: *IEEE Transactions on Computational Intelligence and AI in Games* **6** (2), 228 –242.
URL: http://search.ebscohost.com/login.aspx?direct=true&AuthType=athens&db=inh&AN=14379637&site=eds-live.

Meehan, J. R. (1977) 'TALE-SPIN, An Interactive Program that Writes Stories'. In: *In Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, 91–98.

Monge, P. R. and Contractor, N. S. (2003) *Theories of communication networks*. Oxford University Press New York.

Portet, F. et al. (2009) 'Automatic generation of textual summaries from neonatal intensive care data'.

In: *Artificial Intelligence* **173** (7–8), 789 –816. ISSN: 0004-3702.
DOI: 10.1016/j.artint.2008.12.002.
URL: http://www.sciencedirect.com/science/article/pii/S0004370208002117.

Reiter, E. and Belz, A. (2009) 'An Investigation into the Validity of Some Metrics for Automatically Evaluating Natural Language Generation Systems'.
In: *Comput. Linguist.* **35** (4), 529–558. ISSN: 0891-2017.
DOI: 10.1162/coli.2009.35.4.35405.
URL: http://dx.doi.org/10.1162/coli.2009.35.4.35405.

Reiter, E. and Dale, R. (1997) 'Building Applied Natural Language Generation Systems'.
In: *Nat. Lang. Eng.* **3** (1), 57–87. ISSN: 1351-3249. DOI: 10.1017/S1351324997001502.
URL: http://dx.doi.org/10.1017/S1351324997001502.

Reiter, E. and Sripada, S. (2002) 'Human Variation and Lexical Choice'.
In: *Computational Linguistics* **28** (4), 545–553. ISSN: 0891-2017.
DOI: 10.1162/089120102762671981.
URL: http://dx.doi.org/10.1162/089120102762671981.

Rico, J. M. (2013) *Operation Rheinübung*. manual. Accessed 2014.01.08. KBismarck.com.
URL: http://www.kbismarck.com/operheini.html.

Smith, C. (2003) *Modes of Discourse: The Local Structure of Texts*.
Cambridge Studies in Linguistics. Cambridge University Press. ISBN: 9781139435413.
URL: http://books.google.co.uk/books?id=okPPSq1G10QC.

Sripada, S. G. et al. (2001) 'A Two-stage Model for Content Determination'. In: *Proceedings of the 8th European Workshop on Natural Language Generation - Volume 8*.
EWNLG '01. Association for Computational Linguistics. Toulouse, France, 1–8.
DOI: 10.3115/1117840.1117842.

Sripada, S. G. et al. (2014) 'A Case Study: NLG meeting Weather Industry Demand for Quality and Quantity of Textual Weather Forecasts'. In: *INLG 2014*, 1.

Sternberg, M. (2004) 'Telling in Time (I): Chronology and narrative theory'.
In: *Narrative Theory: Special topics*. Ed. by Bal, M.
Critical concepts in literary and cultural studies. Routledge, 93–137.
ISBN: 9780415316590.

— (2010) 'Narrativity: From Objectivist to Functional Paradigm.'
In: *Poetics Today* **31** (3), 507 –659. ISSN: 03335372.
URL: http://search.ebscohost.com/login.aspx?direct=true&AuthType=athens&db=a9h&AN=61353914&site=eds-live.

Tolkien, J. R. R. (1966) *The Lord of the Rings, by J. R. R. Tolkien*. English. 2nd.
Allen & Unwin London.

Wark, S and Lambert, D (2007) 'Presenting the story behind the data: enhancing situation awareness using multimedia narrative'. In: *Proceedings of MILCOM*.
IEEE. Orlando, Florida. DOI: 10.1109/MILCOM.2007.4455058.

Wikipedia (2014a) *Battle of the Denmark Strait — Wikipedia, The Free Encyclopedia*.
Accessed 2014.8.12. URL: http://en.wikipedia.org/w/index.php?title=Battle_of_the_Denmark_Strait&oldid=619880974.

— (2014b) *Last battle of the battleship Bismarck — Wikipedia, The Free Encyclopedia*.
Accessed 2014.8.12. URL: http://en.wikipedia.org/w/index.php?title=Last_battle_of_the_battleship_Bismarck&oldid=619880947.

— (2014c) *Operation Rheinübung — Wikipedia, The Free Encyclopedia*. Accessed 2014.8.12. URL: http://en.wikipedia.org/w/index.php?title=Operation_Rhein%C3%BCbung&oldid=610961916.

Appendix A: Event-set File Format

The set of events is contained in a text file one event per line. The file's syntax in extended Backus-Naur form is:

```
lines = { line };
line = event-name, "," , time, "," , element-list, "," ,
       parameter-list, "," , significance, ?EOL?;
event-name = term;
element-list = list;
parameter-list = list;
significance = number;
time = number;
number = ?Floating point number?;
list = "(", term, { "," , term }, ")";
term = identifier-character, { "," , identifier-character };
identifier-character = ? Letter or digit excluding "," and ")" ?;
```

For example:

```
prinofw_fire,691,(Princeofw,Bismarck,),(The Prince of Wales fired both of its
   forward turrets at Bismarck),7
german_fire,693,(Bismarck,Prinz_Eugen,Hood,),(The Bismarck and Prinz Eugen
   concentrate their fire on HMS Hood),7
hood_hit,694,(Prinz_Eugen,Hood,),(The Prinz Eugen hits the Hood near the hangar
   by the after mast with its 4th salvo),6
hood_fire,694·1,(Hood,),(A fire starts on the Hood near the portside ammunition
   lift),6
lutjen_orders,697,(Prinz_Eugen,Lutjen,),(Lutjens commands the Prinz Eugen to
   change its target to the Prince of Wales),5
course_change,698,(Hood,),(The Hood changed course by 20 degrees to port to
   bring the after turrets into action),5
```

Note that although events are shown (above) wrapped into multiple lines, there is one event per line in the text file.

The algorithm described in this report operates only on the time, element list and significance of each event. The name of the event is only used to provide a unique identifier for each event to aid in interpreting the operation and performance of algorithm. The parameter list is a place holder which could be used to contain information used by later stages of the IMMP. Since the curent work does not implement these later stages, the parameter list was used to contain a human authored text snippet (from the source material) describing the event.

UNCLASSIFIED

DEFENCE SCIENCE AND TECHNOLOGY GROUP
DOCUMENT CONTROL DATA

1. DLM/CAVEAT (OF DOCUMENT)

2. TITLE

Event Sequencing for Situation Narratives

3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)

Document	(U)
Title	(U)
Abstract	(U)

4. AUTHORS

Ian Dall and Bradley Donnelly

5. CORPORATE AUTHOR

Defence Science and Technology Group
506 Lorimer St,
Fishermans Bend, Victoria 3207, Australia

6a. DST Group NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DST-Group–TR–3351	AR-016-821	Technical Report	April, 2017

8. Objective ID	9. TASK NUMBER	10. TASK SPONSOR
AV14709219	LRR 07/429	

13. DST Group Publications Repository

http://dspace.dsto.defence.gov.au/dspace/

14. RELEASE AUTHORITY

Chief, Joint and Operations Analysis Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SOUTH AUSTRALIA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS

Yes

18. RESEARCH LIBRARY THESAURUS

narrative, situation awareness

19. ABSTRACT

Situation Awareness is a critical factor for decision making in complex dynamic environments. Information fusion techniques enable machines to augment human situation awareness, provided the machine representation of the situation can be conveyed to the human operator reliably.
The problem, addressed in this report, is to order a set of events (a plot) in such a way as to result in a narrative which maximises audience engagement, but does not mislead with respect to the temporal relationship of significant events. A novel algorithm is proposed which uses a contrained optimisation approach to achieve an efficient implementation. Measures of *temporal tension*, *continuity* and *centrality* are proposed and used as optimization criteria. Applying different weights to these measures produces narratives with different characteristics.
The algorithm has been implemented and the results of testing using a large historical naval battle (the World War II sinking of the battleship Bismarck) are presented. With appropriate weights the algorithm was shown to produce a plot comparable to human authored text. A number of vignettes are analysed which provides insight into the operation and limitations of the algorithm. In particular, incompleteness in the input data permitted misleading output in some cases. While unsurprising, this highlights the need for careful consideration of the affect of incompleteness and uncertainty on automatic situation assessment aids.
The need for future work has been identified including: refinement of the optimisation measures; broader and more rigorous evaluation; and integration into a complete situation assessment system.

UNCLASSIFIED