

Click: Measuring Relevance between Personal Narratives through Motif Comparison

Kamakshi Duvvuru
Stanford University
kduvvuru@stanford.edu

ABSTRACT

This paper presents a new approach for measuring relevance between personal narratives, based on the notion of shared motifs. In this approach, 50 personal narratives were parsed to build a dictionary of motifs mapped to their associated words. Then this trained dictionary was used to automatically measure the relevance of any two narratives by comparing the motifs they contain. Finally, the effectiveness of this measure was evaluated by comparing it to human judgments of relevance for the same narratives. Results showed that similarity in motifs is a strong predictor of narrative relevance, but could be further improved by recognizing directly contrasting motifs as a measure of relevance as well.

Author Keywords

Personal narratives; narrative matching; narrative relevance; narrative similarity; narrative motifs; motif dictionary;

ACM Classification Keywords

I.2.7 Artificial Intelligence: Natural Language Processing – Discourse

INTRODUCTION

Personal narratives are being shared at a rate unlike any other time in history, because of the Internet. In contrast to online interactions that are dominated by 140-character expressions and filtered photos, longer personal narratives represent a pocket of depth and rawness.

Because personal narratives are often seen as the most revealing representations of one's identity, their prominence on the Internet means growing volumes of new, machine-readable insight into individuals. This trend has resulted in sets of data, which, if parsed and mapped for relevance to one another, would have powerful implications for social computing.

However, contemporary models of semantic similarity, which involve structural dimensions, linguistic features, and word frequencies, have been found insufficient for judging sameness

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citations on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the authors/owner(s).

ACM 978-1-4503-2406-9/13/10.
<http://dx.doi.org/10.1145/2508468.2514726>

in narratives. [1]

In particular, they often operate at the keyword or topic level, failing to consider the aspects that separate narrative from expository text or a set of disconnected facts. "This sense of "storiness" has yet to be identified and exploited on a large scale." [2]

This paper presents a new model of relevance for narratives that leverages a narrative's unique aspects, or "storiness." In designing this approach, first this vague concept of storiness was distilled into two concrete elements: experiences, which are the essence of the events in a narrative, and emotions, which are the essence of the reactions to those events.

A specific experience or emotion can be connoted by a large variety of words and phrases, and therefore cannot be detected by looking for the same keyword or phrase between two narratives. Thus, the model presented here draws on the approach of Linguistic Inquiry and Word Count (LIWC), "a text analysis software that calculates the degree to which people use different categories of words across a wide array of texts, including emails, speeches, poems, or transcribed daily speech." [3]

LIWC software uses the approach of a domain dictionary, a mapping of domains (such as health, home, and social) to sets of words that connote or constitute each of them, respectively. While the domains themselves are too broad to represent specific emotions or experiences, we adopt the concept of a dictionary for mapping experiences and emotions to their respective sets of related words.

Representing a particular emotion or experience through a set of associated words (i.e., abandonment = {orphaned, alone, left behind, did not care for, lost}) rather than a particular keyword or phrase allows for us to capture the elements of a narrative with the complexity they warrant. For example, one narrative might express a sense of abandonment through the phrases "left behind" and "lost," while another might do so through "did not care for" and "orphaned". A simple keyword frequency approach, evaluating each text as a series of disconnected, discrete terms, would determine that these two narratives are dissimilar, whereas a dictionary approach such as ours, evaluating each text as a set of cohesive themes, captures the common motif woven between the lines of each of these narratives.

This paper describes the development of this motif dictionary approach and an evaluation of it that suggests it is a robust and representative model for comparing personal narratives for relevance.

METHODOLOGY

Corpus of Personal Narratives

210 personal narratives were collected from a web service called The Listserve in which one person a day wins the chance to write to the remaining ~25,000 subscribers. This was chosen as the corpus because the nature of the opportunity to send one message to thousands of strangers fosters stories that are particularly central to and revealing of the writer's personality. This allows for the relevance determined by this approach to have more significant implications in the social computing realm.

The following is a sample narrative from the corpus:

"Recently I almost died.

My esophagus tore open when I was cycling the Alpine Dam/Seven Sisters loop, just north of San Francisco. My chest blew up like a balloon. I went to the hospital that evening. They said I had about a 60% shot at making it more than a day or two. I was alone in the hospital all night. My extended family lives far away, and my wife couldn't leave work.

That night I realized I'd never confronted what it meant to die.

I'm a young guy, early 30s. At an intellectual level I know I won't live forever. And out of vanity and ego, I've always told myself that I'll die without fear and without regrets. It turns out that looming death makes that kind of knowledge cold comfort, and reveals the flimsiness of those self-perceptions. I was afraid I'd die alone and I regretted that I hadn't done more with my life. And most of all, I realized I didn't know how to die well. If I had gone that night, I would have gone out gasping and terrified, not with grace, love, and simplicity.

In the end I lived (clearly!). But the question remained: how does one confront death honestly?

Here's the best answer I've found so far. Apologies to Thich Nhat Hanh, for so roughly approximating his thoughts on the matter. Here goes.

The ocean is full of waves that go up and go down. Some are towering and immensely powerful, others small and gentle. They start far out at sea, and then they crash on the beach and are gone. At the same time, a wave is the water. And the water is the wave. You can't separate the two. When the wave crashes, it becomes water again...which it always was. The beginning and end of a wave are like a person's birth and death. We are ourselves, and we are also made up of everything else: other people, the earth, plants, sun, the sky, and all the elements. We are and have always been inseparable. And we begin, and ultimately, we die. And when we die we become what we have always been: everything else.

And while we celebrate the birth of a wave (so exciting to see one coming!), we don't mourn its death. Because we know that wave is the water, and the water still lives, and will live for longer than we can imagine. We're like the waves. We're born and we die, and at the same time we're not born and we don't die.

That's it. Reach out anytime! Always happy to meet fellow "waves" on this trip across the ocean. Especially if those waves like road biking, craft beer, and video games. :) [4]

Developing the Approach

To understand which features are strong predictors of relevance between personal narratives, both keyword and topic based approaches of semantic similarity were tested for effectiveness in measuring relevance between narratives. The results obtained from these approaches can be summarized as follows:

1. Pure term frequencies do not correlate well with relevance between narratives, because even when two narratives share the same themes, they often have only one or two exact terms in common. Therefore, this approach is too tight of a similarity comparison.
2. Approaches such as the LIWC dictionary that compare broad categories of terms (i.e. humans, money, leisure) also do not correlate well with relevance between narratives. In this case, the dictionary approach interprets themes too broadly, marking two stories that are about widely varied topics as similar.

These results suggest that personal narratives have unique features that prevent them from being compared using traditional methods for identifying semantic similarity. To address this problem, a more tailored approach, which leverages the key elements of narratives and lies in between the two polar approaches, is proposed.

These key elements are two categories of motifs: 1) personal experiences and 2) emotions. These motifs, too specific to be represented as domains and too broad to be represented by a specific term frequency, make up the attributes of a narrative that are compared in this approach to measure relevance.

Therefore, the approach presented here is to build a dictionary like LIWC with narrower categories, or motifs (personal experiences and emotions), and compare the frequency vectors of these categories in each narrative to each other to measure relevance.

Training

50 of the 210 personal narratives in the corpus were set aside for training while the remaining 160 were used for testing the approach. Each of the 50 training narratives was manually parsed, building the dictionary of motifs with their respective growing categories of related words. An example portion of the dictionary is featured in *Figure 1*.

Testing & Evaluation

We developed the following algorithm for comparing the motifs in any two personal narratives for relevance to each other: 1) build vector models of the motif frequencies in each narrative based on the motif dictionary and 2) compare the vector models with cosine similarity to output a relevance score between 0 (least relevant) and 1 (most relevant).

Resigned with Life	At Crossroads	Student	Athletic	Social Justice	Connection/People	Existential Crisis	Art	Current Events/Politics	Comedy	Theater
had it with	I have no idea	student	olympics	social worker	people	infinite	artist	president	joke	troupe
too much to handle	lorn	university	marathon	inner city	love	beyond	art	relations	what do you see	masks
done	I dont know	graduate	triathlon	kids	enchanting	capacity	beauty	industry	knock knock	character
cant	I do not know	learn	yoga	teach for america	wisdom	understand	passion	investigation	funny	venue
can't	I don't know	learned	climb	help	lives	universe	human being	attack	comedy	perform
Im not	become	learned	mountain	passion	stories	complexity	image	hack	laugh	travel
I'm not	career	learning	meditation	NGO	advice	human	poetry	governor	standup	style
disappoint*	student	deciding	ultimate frisbee	volunteer	world	consciousness	speak	committee	stand up	extemporane
over	do with my life	decision	frisbee	opportunity	mind	outside	reflection	platform	stand-up	script
die	do after	decided		change	hope	understanding	inner	fiscal	comedian	actor
dying	decision	decide		justice	human	limitations	see	responsibility	perform	story
dead	crossroads	decide		marginalized	consciousness	human being	hobby	taxes		performance
tiring	change	school		community	link	beauty	existence	government		money
tire	plan	college		injustice	neighbor	immense	connecting	money		proceeds
tired	deadline	spend		lives	global	passion	story	years		charity
confusing		night		matter	community	know	painting	people		show
confused		months		loved ones	young	want	creative	presidential		costume
confuse		champagne		loss	old	hunger	design	campaign		props
lonely		drinking		event	race	inner	illustrate	politics		reimagining
alone		friends		issue	colour	planets		future		original
not working		PhD		family	creed	galaxies		organization		pieces
trying		Chemistry		nonprofit	rich	vast		institute		written
I just cant		pursue		social	poor	infinity		train		director
stick		liberal arts		better place	same	larger		recruit		write
comfortable		small		world	earthing	inescapable		expand		family
not		dissertation		better	eat	end		innovate		rehearsal
desperate		major		brighten	love	imagine		conference		production
sucked		party		foundation	healthy	ourselves		gathering		producer

Figure 1: Excerpt from Motif Dictionary

Using this algorithm, we compared each of 160 personal narratives against one reference narrative, and ranked them from smallest to largest relevance scores.

In order to evaluate the effectiveness of the algorithm in ranking relevance between personal narratives, we generated a survey to collect human ranking judgments for the same narratives to serve as a benchmark for the algorithm.

However, the number of narratives that could be evaluated by a survey participant was limited by time and effort to about 6 narratives (as opposed to the 160 from the algorithm).

While this is number is a seemingly non-representative sample of the larger dataset, two measures were taken to increase the validity of the small sample:

1. The 6 narratives to be evaluated were split between two evaluation tasks so that all the results would not revolve solely around one reference narrative. So, two narratives were chosen as the reference narratives, each of which would be evaluated against three narratives for relevance.
2. The three narratives to be evaluated against each reference narrative were chosen by running the algorithm with respect to that reference narrative, and randomly choosing one narrative from each third of the results (one each from the 33rd, 66th, and 99th percentiles of the algorithm's 160 ranked narratives). This most closely simulates the algorithm's ranking of three narratives against one reference narrative.

In summary, the survey consisted of two evaluation tasks. Each of the tasks featured one reference narrative and three narratives (which span the range from least similar to most similar according to the algorithm's rankings) that must be ranked for relevance to the reference narrative. At the end of

each evaluation task, the survey also asked participants the degree to which each aspect of the personal narratives—topics, author's personality, author's interests, or other—influenced the participants' relevance ranking during that evaluation. 27 participants, 12 males and 15 females between the ages of 20 and 30, were recruited to complete the survey.

RESULTS

Below are results from the two evaluation tasks of the survey, representing algorithm vs. human rankings of relevance for each set of narratives in Charts 1 and 2 and factors influencing human rankings in Charts 3 and 4.

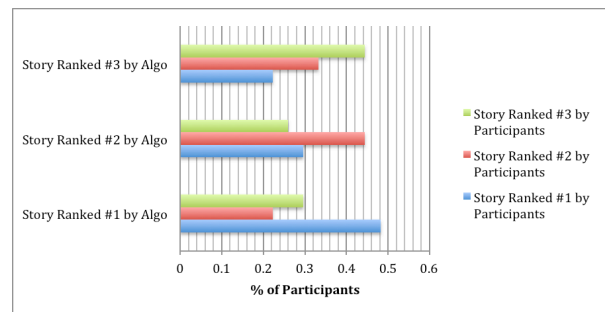


Chart 1: Algorithm vs Participant Ranking (Evaluation Task 1)

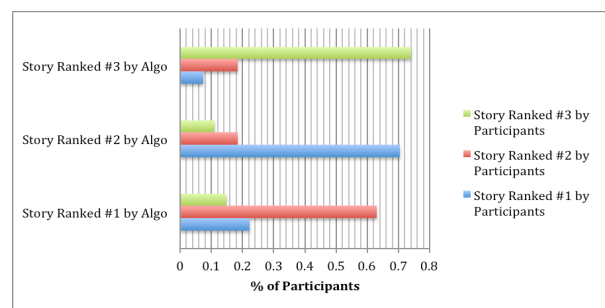


Chart 2: Algorithm vs Participant Ranking (Evaluation Task 2)

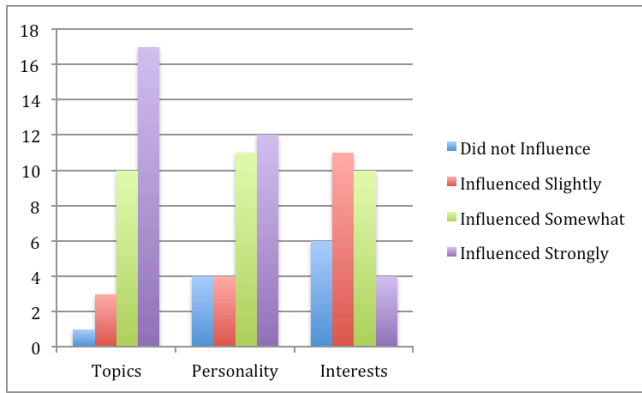


Chart 3: Influential Factors in Relevance Ranking (Evaluation Task 1)

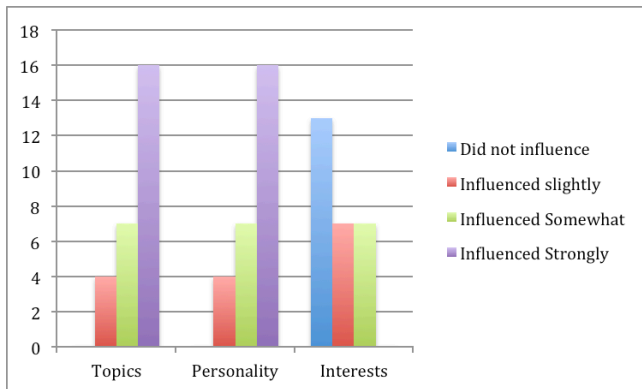


Chart 4: Influential Factors in Relevance Ranking (Evaluation Task 2)

DISCUSSION

Chart 1 demonstrates that the participants' relevance rankings of the narratives align with the algorithm's rankings in evaluation task 1. That is, for each of the three stories, the largest group of participants assigned the story the same relevance rank as the algorithm did.

However, this finding is complicated by the results of evaluation task 2, featured in *Chart 2*. As the results show, for the story that the algorithm ranks #3, the largest portion of participants also rank the story #3. However, for the stories that the algorithm ranks #1 and #2, the largest portions of participants give a rank of #2 and #1, respectively.

As *Chart 3* and *Chart 4* demonstrate, the participants' rankings in both evaluation tasks were "strongly influenced" by the same two factors (topics and personality). Therefore, a deeper investigation is required in order to understand where this inconsistency in the algorithm's ranking effectiveness between evaluation 1 and 2.

The next set of data that could offer this explanation is the list of participants' open-ended responses about their ranking processes.

The following are the set of qualitative responses from each evaluation task, in response to the question, "Anything other

than the above factors (topics, personality, interests expressed in the narrative) that influenced your ranking?"

Evaluation Task 1: Open Responses

The tone of the passage was a big influence in how I perceived each passage, and thus how closely I judged it to be to the bolded passage.

All of these had to do with death, so it was hard to tell if topics/experiences was a factor.

First, I considered irritability. It seems like bolded and 3 would be really annoyed at each other - bolded =serious; 3 = lighthearted. Then, beliefs - both bolded & 1 had a turn around experience regarding death. Finally, interests.

Life is precious

How serious I felt they were when they wrote the essay. The bold essay kind of undermined its own seriousness when he said he wanted to meet fellow waves in the ocean. At no point in his earlier paragraphs did I realize he wanted to meet people. And I judged them all for his likes and dislikes.

Age (or implied age) as a proxy for life stage and experience

Whether I considered them to be a "good writer," strangely enough

The first two seemed more positive/constructive than the third.

Gender and close brush with death; age and childlessness

Fairness of the situation each of these people were in.

I thought the ones that I put as 1 and 2 were looking to also listen to other people (as opposed to just talk to them) but the one I put 3rd just looked like she just wanted to give advice, so I generally am averse to that.

Shared values

I thought it would be meaningful to befriend the 52 yr old because he has slightly different views on death than the bolder author. He's afraid of dying and I think that fear is something the bolded author is trying to find an answer to. Whether they arrive at the conclusion or not, I think they would have meaningful discussions and seek each other out.

Evaluation Task 2: Open Responses

The point of view influenced my rankings for this one. To me, the bolded story speaks a lot about the writer's personal struggle, accepting yourself, and figuring out how to move forward. But even though the next story has a similar topic, it starts off with a very long poem that's written FOR other people rather than describing the writer's own experiences, so it just takes away from the connection and feels impersonal. I almost feel the story would've been more powerful if the "I am just a foolish young person" paragraph came at the beginning rather than at the end, and I would've ranked it at most relevant rather than second most relevant.

I ranked the first one because this person was offering to give something that the narrator was asking for while the others may

not have been able to cheer up the narrator as well. It's not only about two people being similar but also having a give and take that fits.

I thought the first person was pretty annoying and self-indulgent. So I ranked them best on the POV of the first person, from who he would like most to who he would like/relate to least. #1 seemed likely to permit the bold essayist to be self-indulgent, the second would tolerate it a little and the third has a different philosophy from the bold one (but not the second-ranked one).

Whether or not their response seemed targeted towards the bolded author or not. While two of them asked for feedback/thoughts or had an invitation to continue a conversation, they seemed somewhat self-centered and focused on their own experiences w/o really addressing the bolded author's expressed emotions/points.

I put the 38 year old at the end because he seems to be in the same place as the 21 year old. Although they seem to be experiencing the same emotions and frustration with the world, I don't think it's quite as inspiring for the 21 year old to see a 38 year old in the same place. I think it could be quite discouraging – that the 21 year old might think, "Wow. I could still feel like I feel now in 17 years." Even though the bolded author asked for poetry, I chose to put the poem in second place because the language, though uplifting, felt forced instead of genuine. Sure, it may have been well-intended, but it didn't feel informed. It felt cookie cutter - like what you're supposed to say. I think the bolded author would have seen through that.

Level of angst

These responses provide an important insight: participants rank two narratives as relevant not only when they feel these narratives are analogous, but also if they feel that one is an apt response to the other.

The latter is mentioned much more prominently as the reasoning behind ranks during evaluation task 2. Comparing the set of narratives presented evaluation task 1 and 2, we can glean that this was likely a result of the following:

- Narratives in the first evaluation task all broadly address the topic of life and death, but no two narratives can actually be seen as the opposite view or response of another. Therefore, these stories lend themselves more to a ranking based on similarity.
- Narratives in the second evaluation task, however, contain pairs that are precise responses or opposites of one another as well as pairs that are analogous to one another. When such “perfect opposites” exist in the mix, participants seemed to actually find these pairs of opposites (complementary views on the same topic) more relevant to one another than pairs of matches (same views on the same topic).

Specifically, the reference narrative in the second evaluation task was a young adult's description of a low and hopeless point in their lives. Most participants chose to rank a narrative

about maintaining hope during rough times higher in relevance than a narrative that shares the same hopeless perspective as the reference narrative.

Ultimately, the inconsistency in the effectiveness of the algorithm between evaluation tasks 1 and 2 can be attributed to the fact that the algorithm was designed with a narrow definition of “relevance,” overlooking the idea that contrast can be just as strong of an indicator of relevance as similarity.

CONCLUSIONS AND FUTURE WORK

The fact that the algorithm's definition of “relevance” was narrower than that of the participants creates a misalignment of rankings between the algorithm and human rankings in the second evaluation task, when contrasting motifs become a stronger predictor of relevance.

However, based on the alignment of the results in the first evaluation task, in which similar motifs are the stronger predictor of relevance, it is apparent that, configured with the appropriate indicators of relevance in mind, the algorithm's model is effective in recognizing similarity between narratives at a thematic level.

Therefore, the motif dictionary model has allowed for an improved measure of narrative relevance by comparing narratives with respect to their cohesive elements or “storiness” that are unique to them, rather than the attributes common to all text such as structural components or sets of keywords and phrases.

In future iterations, the motif dictionary needs to be reconfigured to account for the importance of contrast as an indicator of relevance, i.e., in a way that recognizes complementary views of the same topic as a single motif. This will account for the discrepancies between human and algorithm definitions of relevance that became apparent during this study.

Further, in order to gain more significant results, future iterations should strive for the number of narratives ranked by participants to more closely match the number of narratives ranked by the algorithm.

REFERENCES

1. Fisseni, B., & Löwe, B. (n.d.). *Which Dimensions of Narratives are Relevant for Human Judgments of Story Equivalence?* (pp. 114-117). Amsterdam: Institute for Logic, Language and Computation.
2. Elson, D. (2012). Literary Social Networks. In *Modeling Narrative Discourse* (pp. 1-42). Columbia University.
3. Development and Psychometric Properties of LIWC2007. (n.d.). Retrieved December 27, 2014, from <http://www.liwc.net/howliwcworks.php#index2>
4. Waves. (2014, April 17). Retrieved December 27, 2014, from <http://thelistserve.com/>

5. Krakauer, C. (2012). *Story retrieval and comparison using concept patterns*. Massachusetts: Massachusetts Institute of Technology.

6. Kypridemou, E., & Michael, L. (2013). *Krakauer, C. (2012). Story retrieval and comparison using concept patterns*. Massachusetts: Massachusetts Institute of Technology. Nicosia: Open University of Cyprus.