Automatic Bill Recommendation for Statehouse Journalists

Michelle Perkonigg[1(✉)], Foaad Khosmood[2], and Christian Gütl[1]

[1] Graz University of Technology, Rechbauerstraße 12, 8010 Graz, Austria
michelle.perkonigg@student.tugraz.at, c.guetl@tugraz.at

[2] Institute for Advanced Technology and Public Policy, California Polytechnic State University, San Luis Obispo, CA 93407, USA
foaad@calpoly.edu
https://iatpp.calpoly.edu/

Abstract. AI4Reporters is a project designed to produce automated electronic tip sheets for news reporters covering the statehouses (state level legislatures) in the United States. The project aims to capture the most important information that occurred in a bill discussion to allow reporters to quickly decide if they want to pursue a story on the subject. In this paper, we present, discuss and evaluate a module for the tip sheets that is designed to recommend additional bills to investigate for the reporter that receives the tip sheet. Similar in concept to movie recommendations, this module is designed to find other bills with their own meetings and discussions, that are most relevant to the discussion captured in the given tip sheet. Specifically we present similarity algorithms along three dimensions that our investigation suggests are distinct reasons for journalists to be interested in a recommendation. These include similarity in content, individuals or geographical locations. We validate the system by fielding a user study of 29 subjects for hour-long surveys resulting in 870 decisions being captured. We find that between 63.4% and 82.8% of the human selections are in agreement with our system's recommendations.

Keywords: digital government · legislatures · bill recommendation · artificial intelligence

1 Introduction and Motivation

AI4Reporters [25] is a project aiming to create AI-powered, automated tip sheets generated for reporters that are otherwise unable to cover the legislature in person. A kind of algorithmic journalism [23], AI4Reporters processes the transcript and video of a legislative hearing and then generates interesting facts, anomalies (such as unusual voting patterns), pull quotes, speaker lists, backgrounders and other useful features in form of a web-accessible interactive tip sheet [25] or a full news story [27]. Most tip sheets are generated per bill discussion (a subdivision of a committee hearing focused on discussing and voting on a single bill). The

Published by Springer Nature Switzerland AG 2023
I. Lindgren et al. (Eds.): EGOV 2023, LNCS 14130, pp. 128–143, 2023.
https://doi.org/10.1007/978-3-031-41138-0_9

idea is this information could provide a tip for a reporter to help them make a decision to pursue a story on the subject.

If the reporter does decide to investigate further, the tip sheet provides many references to useful background information, each linked to verifiable, primary sources for complete transparency and traceability. One of the main elements that is always necessary for such investigations is related or similar bills that are either going through the legislative cycle or have already completed it. In this paper, we present, discuss and evaluate a bill recommendation module for the tip sheets designed to surface a few relevant bills for the reporter to consider. The bulk of our work described here is development and evaluation of an algorithm for this recommendation system. The proposed algorithm is a first suggestion to be adapted by community due to the novelty of the whole system in the application domain.

1.1 Motivation

In this section we present the motivation for the parent project and also for the present work which is a recommendation module for tip sheets.

AI4Reporters. Unlike the US Congress, European Parliament and numerous national legislatures, written proceedings are not officially produced or maintained by US state governments, effectively cutting off meaningful access to vast majority of citizens and researchers [8]. While the governments do publish bill titles, bill texts, committee memberships, and vote outcomes, there is a considerable gap in knowledge in the absence of written, searchable records of spoken language.

Until about the first decade of the twenty first century, the aforementioned gap was mostly addressed in the form of news reporting. While most ordinary citizens in a state like California, could not travel to Sacramento and would not have direct access to legislative information, they would still get the highlights from their hometown newspaper, radio station or TV station. A vibrant cadre of journalists representing many cities, towns and rural areas in the state, used to flood the buildings of the California legislature, be present at hearings, and make sure developments important to their readership would be covered.

A number of factors disrupted the local news economy which in the past twenty years resulted in severe decline in state and local reporting. Among them are competition from internet news sources and media corporate consolidation leading to many traditional regional news media organizations being purchased by large corporations that prioritize national over local coverage. Analysis of the factors leading to the changing media landscape and the reasons for them are beyond the scope of this paper. We only emphasize the present reality of severely diminished news coverage at the statehouse [19,25,39].

The notable absence of media covering the legislature can have some devastating consequences for citizens in a democratic society, even at the state level. Some of the most important legislation with global impact is discussed and debated there. California alone is on the verge of becoming the world's fourth largest economy with $3.63 Trillion GDP [44]. Not only are citizens deprived of valuable information, but they have decreasing opportunity to hold lawmakers accountable for their actions and statements. Meanwhile, well-resourced and powerful interests who can afford to hire lobbyists have better access and more influence with the legislatures.

Thus the overarching motivation of AI4Reporters is to strengthen local and state media and to help increase accountability and transparency by democratizing access to legislative proceedings [25].

1.2 Recommender Module

When reporters use electronic tip sheets to keep informed on the events of a committee bill discussion, they will at some point decide if there is reason to pursue a news story with a more complete explanation. In order to prepare for that story, or even when trying to decide on writing it, the reporters need to examine other, similar, discussions to be able to get a better context. The recommender module is designed to give them a quick list of one to three references for examination.

Reporters can of course dive deeper and familiarise themselves with a much larger set of bills for their background investigation. They may decide to read every single bill passed in that committee or all the previous bills authored by a certain individual. We aim to provide only the first step, a quick glance on what else could be relevant.

One of the main questions that arises early in this work is "by what criteria should relevance be measured"? Based on discussions with area experts on the project and observations of the state legislative proceedings in California, we identify three main dimensions to this notion of "relevance": people, locations and issues. These are based on typical assignments for a reporter. For example, a local reporter may be primarily interested in their representative or bills mentioning their locale and thus would find recommendations of bills involving the same individuals or geographical entities compelling. Similarly, a reporter may be following an important issue and thus would be open to recommendations of other bills discussing similar issues.

We further present three scoring systems as means to automatically measure each dimension, breaking down each score into components derivable from the given corpus. Our hypothesis is that bills selected based on our system will match user expectations of a good recommendation to a significant degree. We test the system with a user study and generally find that study subjects agree with our algorithms in each of the three areas by majorities of 63.4% (locations), 75.2% (people) and 82.8% (content). See Fig. 4.

2 Background and Related Work

In the domain of legislature and legislative proceedings there is a broad range of different research directions to be considered from prediction of votes on legislators [7], over the prediction of bill survival [45] to supporting the drafting phase of a bill [1], to fully producing articles automatically about a hearing [27].

Due to this kind of support reporters can spent less time crawling through the huge amount of available data and defining relevant facts [25]. Focusing on this data, the documents and the contained language have to be processed which requires the field of Natural Language Processing (NLP) to come into action, along with machine learning and artificial intelligence. The aim is to give computing units the ability to communicate in a human manner, such that natural language can be processed and analyzed correctly and therefore enable a human-like response or behaviour involving semantic appropriateness [4].

NLP pipelines often involve several preliminary or pre-processing stages, such as lemmatization, stemming, tokenization, part-of-speech tagging and entity recognition combined with document clustering [28], semantic analysis [37], supervised machine learning and many more in a broad application area.

2.1 Legislative Analysis

Researchers have explored predicting votes in legislatures. [22] presented a method for that prediction using an ideal point topic model. Therefore historical legislative voting data and bill texts were used to conduct topic modeling on the bill text and determine an ideal point for every legislator to finally calculate the prediction using the model. [24] focused on predicting votes in the U.S. on topic level, also based on using an ideal point estimation for every topic. [9] did it at the state level (California).

Another direction of research is to predict bill survival implying the likelihood of a bill to become a law [7,45].

Another area of interest is the support of individuals in different phases of the legislative proceedings. Those supporting methods can be performed for better understanding of legislation. Within this field, [1] presented a compliance assessment tool for EU legislation that delivers descriptions of legal terms, soft-obligations, exceptions and related legislation to a legislation of interest.

Another supporting system is Quick Check introduced by [43]. Quick Check recommends relevant cases to a legal issue given by a user by applying different methodologies for extracting document structure, determining potentially relevant cases and ranking to present the most relevant cases.

Still in the area of supporting and providing legislative data, [32] shifted their focus on the storage of this information, suggesting, based on the Belgian legislature, approaches for process automaton to improve timeliness and availability of legislative data.

2.2 Digital Democracy and AI4Reporters

Digital Democracy [8], a project launched by The Institute for Advanced Technology and Public Policy at California Polytechnic State University aims at filling the gap of providing valuable and comprehensible information for citizens as mentioned in the citation above.

One of the main challenges in government transparency in the United States is the availability of proceedings at the state legislatures. US state governments are republics with very similar structure to the federal government. But compared to the national legislature, the state legislatures, such as those in California and Texas, are less studied and less transparent. For example there are no official transcripts of discussions in US state legislatures [36].

The AI4Reporters project, as the title of the project already indicates, uses artificial intelligence that processes data from different sources amongst which is the legislative database populated by Digital Democracy, extracts facts and finally shows it in a readable and well structured way, such that reporters can use this information for their report [25].

As part of the quality legislative database are bill texts, which are the formulated ideas that can become law. This type of text follows a simple shape as demonstrated in Fig. 1 [13]. The parts included in the database of the Digital Democracy project are:

– **The bill ID** is a unique identification for the bill in the session year. It is composed by the type (AB - Assembly Bill, SB - Senate Bill, etc.) and a unique number for that session year.
– **The bill title** gives a short statement of what the bill is about.
– **The bill author** lists all authors and co-authors of the bill.
– **The bill status** describes the current status of the bill. This can either be proposed, introduced, amended assembly, amended senate, enrolled or chaptered.
– **The bill digest** presents a short summary of the bill.
– **The bill text** contains detailed information on the bill content.

The length of such bills can vary tremendously, starting with a small bill where only few sentences are necessary for description (see [11]) going up to bills that consists of several pages (see [12]), that outline and explain the bill, its limitations and its influences in detail.

2.3 Recommender Systems

Nowadays recommendation systems are widely used. For example, Amazon recommending books based on shared interest with other users or Netflix recommending movies and series by predicting ratings for a movie or series [38]. The general problem faced by this systems is the pure overload of information that is still increasing with time. Therefore, limited and carefully selected potentially interesting information is presented to the user based on different underlying recommendation techniques [15].

Basically recommender systems are divided into the two most common categories: content-based and collaborative filtering. Those two are often extended by other typical categories, some of them listed and explained below.

Fig. 1. Preview and structure of a bill (Adapted from: [11]).

Collaborative Filtering. Collaborative filtering generates recommendations by matching the users interests and preferences and with information gathered from other users and their preferences. Therefore, this type of recommendation highlights the necessity of available data implying the dependency on the collaboration of users [5]. Collaborative filtering can further be subdivided into user-based and item-based. User based collaborative filtering searches for similarities between users to recommend new items, while item-based is based on similarity between new items and items contained in the users historical data [30]. Context information can also be employed and integrated into a collaborative filtering technique, allowing the system to provide different recommendations in different situations [17].

Content-Based Filtering. Content-based filtering focuses on historical information of the user (e.g. purchases) and the description of items to generate recommendations [33]. The general approach of content-based filtering is to create a user profile by defining the preferences through analyzing behaviour and personal data. This user profile is then matched with information about the items to filter out the best matching one [30].

Other Kinds of Filtering. Demographic filtering considers demographic data of a user and exploits the attributes of demographic categories of users or items to provide suggestions [6]. [3] analyzed different approaches to profile users. The presented approaches are categorized into unified (mixed, categorical and fuzzy) and isolated (cascaded and single attribute) approaches depending on how the attributes are combined, each of them considering age, gender and occupation as demographic attributes.

In contrast to collaborative and content-based filtering, this approach does not rely on collected historical preference data of the user.

Knowledge-based filtering recommends items with the help of a knowledge base that forms information about users and items. Ontologies are often used to represent information in a structured way, capturing concepts and relations of objects in the ontology [41]. [2] proposed a filtering technique using an ontology that is updated dynamically with new information about users and items.

Hybrid filtering is a combination of different techniques to achieve better results and face each others limitations and problems [10]. [46] verified through the conducted study that a combination of collaborative filtering and demographic filtering (gender, nationality and age) can improve results in the application area of music recommendations.

Application Domains. Recommendation system are employed in many different areas [30] to support users by providing a selection of filtered information. Some of the relevant examples include E-Commerce [20,26], E-Resources focuses on recommending shared content like videos [29], music [16] and documents [43], Digital Libraries [14], E-Government [18,42].

3 Bill Recommendation System

The bill recommendation system is meant to work as a component of those tip sheets which focus on a single bill discussion in the legislature. It produces a number of other similar bills that may be of interest to the reader. Due to the novelty of the application domain, the general concept of the recommendation system is designed based on the insights given by a domain expert and therefore represents a first approach to be adapted in future. Three types of similarity are considered, and thus up to three different recommendations can be made. These are: geographical entities, participating individuals and bill content. For each type of similarity, the system recommends a bill most similar to the one under review. See system architecture in Fig. 2.

3.1 Recommendation Based on Geographical Entities

Recommendations based on extracted geographical entities focus on delivering results that talk about the same geographical location or places and therefore, draw a connection between discussed bills. For this purpose the state of a bill and geographical locations mentioned in the bill text are considered and weighted, such that bills introduced in the same state are prioritized. As soon as geographical entities are extracted, validation thereof is conducted using a geocoding python library [21] to reduce false positively tagged entities. Equation 1 presents the used formula to determine geographical similarity, in which $locations_{BillX}$ is a set of validated geographical entities extracted from the specific Bill X and $state_{BillX}$ holds the US state where Bill X is presented.

$$\begin{aligned} score_{geo} = score_{state} * 0.3 + \frac{|locations_{BillA} \cap locations_{BillB}|}{|locations_{BillA} \cup locations_{BillB}|} * 0.7, \\ \text{with } score_{state} = \begin{cases} 1, & \text{if } state_{BillA} == state_{BillB} \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{1}$$

3.2 Recommendation Based on Individuals

The second recommendation type focuses on participating individuals. The basic idea of this recommendation type relies on the assumption of a shared interest between the reader of a certain bill and the participating groups of people and individuals in this bills' life-cycle. We consider the author of a bill, the speakers during all the bill discussions, the affiliations of the speakers and the organizations mentioned in the bill content. Every extracted entity is validated by checking its entry in the legislative database. Moreover, extracting and validating this data allows us to apply a weighted distribution which we derive experimentally. In the final score shown in Eq. 2, $author_{BillX}$ holds if the two bills share the same author, weighted at 20% of overall importance. The next term, $speakers_{BillX}$, is a measure of mutual speakers present in both bill discussions and is also weighted at 20%. $affiliations_{BillX}$ is similarly a measure of mutual speaker affiliations, weighted at 30%. Finally, $organizations_{BillX}$ represents a measure of mutual organizations mentioned in the discussion. also weighted at 30%.

$$\begin{aligned} score_{individuals} = score_{author} * 0.2 + \frac{|speakers_{BillA} \cap speakers_{BillB}|}{|speakers_{BillA} \cup speakers_{BillB}|} * 0.2 \\ + \frac{|affiliations_{BillA} \cap affiliations_{BillB}|}{|affiliations_{BillA} \cup affiliations_{BillB}|} * 0.3 \\ + \frac{|organizations_{BillA} \cap organizations_{BillB}|}{|organizations_{BillA} \cup organizations_{BillB}|} * 0.3, \\ \text{with } score_{author} = \begin{cases} 1, & \text{if } author_{BillA} == author_{BillB} \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{2}$$

3.3 Recommendation Based on Bill Content

The last type of recommendation is based on the content, outputting a reference bill that shares some similarity with the bill of interest in their contents. Therefore, the similarity graph introduced by [40] is used, which is basically a bidirectional weighted graph connecting words and sentences to each other relying on their relations retrieved from the lexical database WordNet [31]. To determine similarity a bill has to be linked to the graph appropriately. The graph is then exploited in two ways, contributing to two different scores that are then combined to retrieve the final score for content recommendations. First score exploits the structure of the graph in combination with using the Levenshtein Distance [35] to get a rather fast result for determining bill title similarity. Exploiting the structure, without considering the weight and defining a maximum depth allows us to retrieve only semantically close nodes from the graph. In Eq. 3 $nodes_{TitleX}$ represents this set of extracted nodes for the title of Bill X. The second score ($score_{text}$) uses the similarity calculation as proposed by [40] exploiting the linkages and their weights of the graph, by performing a breadth first search to finally get a similarity score for two bill contents. Hereby $Bill\ A \xleftarrow{all\ paths} Bill\ B$ of Eq. 3 refers to the extraction of all paths in the similarity graph going from Bill A to Bill B having a predefined minimum weight and maximum depth. Final score then is composed by the sum of the equally weighted similarity scores of bill title and bill content as shown in Eq. 3.

$$\begin{aligned} score_{content} = \frac{1}{2} * (&\frac{100 - LevenshteinDistance(Title\ A, Title\ B)}{100}) * 0.5 \\ &+ \frac{|nodes_{TitleA} \cap nodes_{TitleB}|}{|nodes_{TitleA} \cup nodes_{TitleB}|} * 0.5) + \frac{1}{2} * score_{text}, \\ &\text{with } score_{text} = \\ min(\alpha, &\frac{\sum Bill\ A \xleftarrow{all\ paths} Bill\ B + \sum Bill\ B \xleftarrow{all\ paths} Bill\ A}{2}), \\ &\text{where } 0 < \alpha \leq 1 \end{aligned} \tag{3}$$

3.4 Development

On implementation side, the bill recommendation system is built in a modular way, allowing easy modification but also extension of new recommendation types.

Since journalism is a rapid business that requires the bill recommendation system to be as efficient as possible, a set of well-defined constraints and steps to enhance system performance are incorporated. Potential recommendations are restricted to bills of the same session year and bills having the same main committee, reducing the number of similarity determinations to enhance performance. Further for content recommendations, an additional constraint is given that the title must share some minimal similarity to be considered for computing the full content similarity. We also use domain specific stop word list excluded from similarity consideration.

The system consists of three components (see Fig. 2). For performance reasons, the component 'Similarity Graph' is generating and storing the graph only once and is not to be updated unless the underlying information for graph construction changes. The component 'Content Scores' runs every night calculating

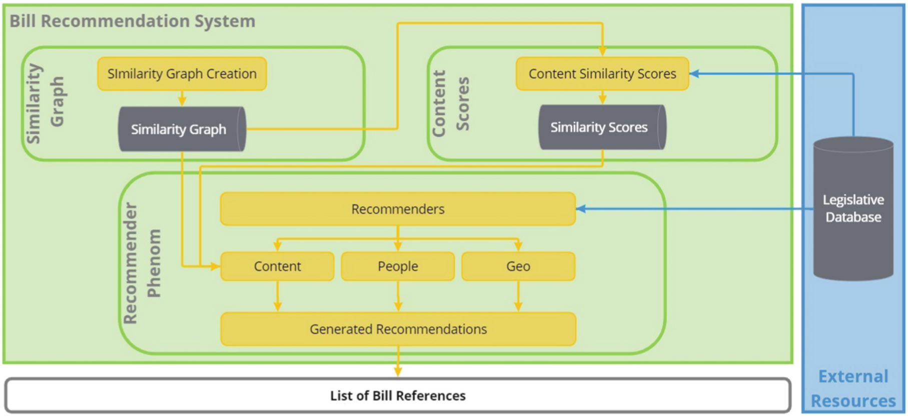


Fig. 2. Conceptual Architecture of the bill recommendation system consisting of three components: Similarity Graph, Content Scores and Recommendation Phenom.

similarity scores for content between bills and lastly the component 'Recommender System' generating scores for a specific bill of interest on demand.

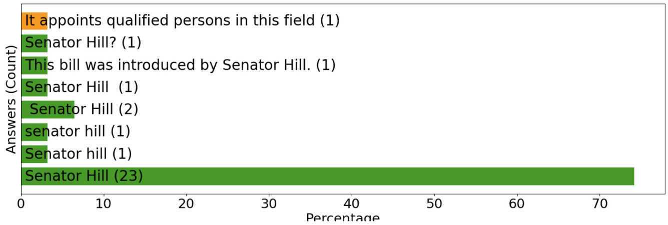
4 User Study

The purpose of the user study is to see if users agree with our systems recommendations and further to see if the underlying recommendation types make intuitive sense to the users.

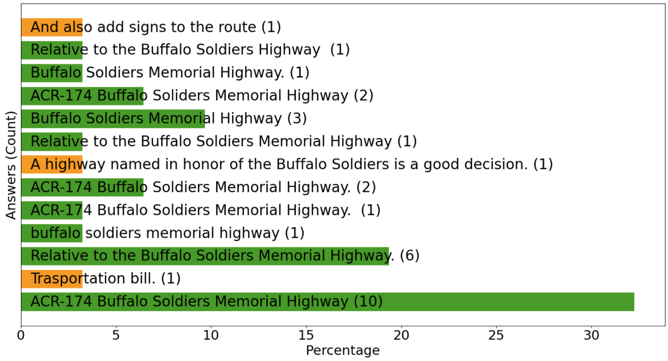
We use the paid online distributed research participant recruitment service Prolific [34], and choose to restrict participants to those located in California who have completed secondary education. The location restriction is realistic for a target audience for such a tool and increases chances the study subjects have familiarity with bill content, locations and individuals.

After a brief opt-in user study informed consent and explanation, the survey consisted of ten pages of content questions. Each page began by asking the user to follow a hyperlink and read a given bill of interest. After this, they were asked to read three other bills as recommendations for someone who was interested in the first bill. The user was asked to read each of those bills, and then to select between two choices of "this a good recommendation" or "this is a bad recommendation", and provide an explanation as to why they answered the way they did.

Furthermore, two control questions were asked on each page to make sure the users were paying attention: they were asked to type in the author of the bill of interest, and its title in free-form response questions. Those survey returns that did not correctly answer these questions were dismissed. Figure 3 provides an overview of the answers to the control questions.



(a) Answers given to question asking for author of a bill.



(b) Answers given to question asking for title of a bill.

Fig. 3. Example accepted answers to the two control questions injected into the user study are highlighted in green, the orange answers were subject to rejection. (Color figure online)

The users had no prior information about how the recommendations were selected. In reality, one of the three recommendations was completely random. Another was generated by our system based on one of the similarity measures. The third recommendation was either random or system generated. In this way, either one out of three or two out of three recommendations were random such that the user couldn't intuit that a majority of recommendations are "good" or "bad" per bill of interest. The survey guaranteed exactly half the overall recommendations shown to a user to be random, and the other half system generated, equally distributed among the three different similarity measures.

The data, bills and transcriptions used for this study are from the Digital Democracy project [8] only considering bills from the California legislature 2015–2018 which aligns with the restriction set for distribution. Running the study resulted in valid answers from 29 out of 31 participants each of them having to rate 30 presented recommendations for 10 bills of interest, consequently 870 decisions were collected. The study is estimated to take an hour, highly depending on the speed of reading of the study participants.

5 Results

Overall, we collected 523 (60.1%) ‘yes’ answers and 347 (39.9%) ‘no’ answers, thus slightly more than 60% of the recommendations were considered as good ones. Diving into more detail, dividing the ratings into their source of recommendation generation (see Fig. 4), it can be seen that for ratings of random recommendation there are mixed opinions among the participants, as expected.

Moving towards the recommendations produced by the proposed bill recommendation system, a positive trend can be seen. Content based recommendations with 82.8% good ratings are outperforming individual based ones having 75.2% and geographical information based ones with 63.4%. For system generated recommendations there’s almost a 2:1 consensus with the user subjects on all three variants.

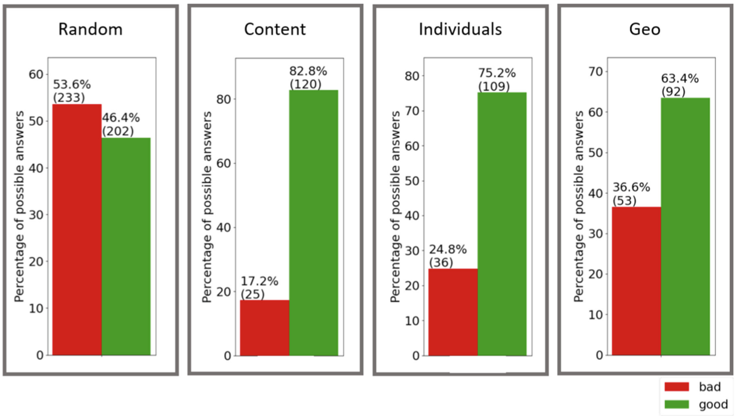


Fig. 4. Breakdown of boolean answers given to the recommendations to be rated in the survey based on their source of generation: random, content based, based on individuals or based on geographical information.

Evaluation of the follow-up short text answers, all of which were coded by the authors, shows that content based recommendations are always recognized as such, due to the answers drawing content-related connections between the two bills. Further investigation into the results conveys the impression that content is the first place to look for similarity of two bills.

With this study we show that the proposed system outputs a relevant but limited set of recommendations with respect to a bill of interest, providing a new source of information to be considered for reporters, with a major advantage that the exact reason for the recommendation can be published alongside it for reader consideration.

6 Conclusion

Proposing a bill recommendation system having a quality legislative database as data source and reporters as target audience is a challenging task, keeping in mind the steady growth of data and the fast business of journalism. With this system a list of related bills for a given bill of interest is presented, that share some sort of similarity, either regarding geographic information, participating individuals or content. Our user study shows the proposed system outperforms randomly presented recommendations significantly. Evaluation of the system-generated recommendations shows that content-based recommendations perform best, followed by recommendations based on involving individuals and those based on extracted geographical information. Further investigation indicates that content is the most important factor when looking for similarity between two bills.

The applied weighting schemes used in our scoring for all three types of recommendation are a first contribution to the community, derived from the provided insights given by experts of the California state legislature active on the AI4Reporters project. Consequently there is the necessity for adaption of those weighting schema, which requires access to a specific group of population to be studied.

An automated tip sheet system for state legislature has never existed before and there are no other systems to benchmark against. We hope to do future field evaluations if and when this proposal is adopted by journalists.

Future work emerging from the findings of the user study point towards more sophisticated and detailed calculation techniques, especially for the individual based and geographical information based recommendation types. This could be in the form of weighting the extracted geographical information by distance, or weighting of speakers based on their speaking time during the bill discussions, but first this needs to be analyzed. However, future work includes analysis for the extension with additional recommendation types, worth to include, while not overpopulating the tip sheet.

References

1. Agarwal, S., Steyskal, S., Antunovic, F., Kirrane, S.: Legislative compliance assessment: framework, model and GDPR instantiation. In: Medina, M., Mitrakas, A., Rannenberg, K., Schweighofer, E., Tsouroulas, N. (eds.) APF 2018. LNCS, vol. 11079, pp. 131–149. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02547-2_8
2. Alaa El-deen Ahmed, R., Fernández-Veiga, M., Gawich, M.: Neural collaborative filtering with ontologies for integrated recommendation systems. Sensors **22**(2) (2022). https://doi.org/10.3390/s22020700. https://www.mdpi.com/1424-8220/22/2/700
3. Al-Shamri, M.Y.H.: User profiling approaches for demographic recommender systems. Knowl.-Based Syst. **100**, 175–187 (2016). https://doi.org/10.1016/j.knosys.2016.03.006. https://www.sciencedirect.com/science/article/pii/S0950705116001192

4. Allen, J.F.: Natural Language Processing, pp. 1218–1222. Wiley, GBR (2003)
5. Ansari, A., Essegaier, S., Kohli, R.: Internet recommendation systems. J. Mark. Res. **37**(3), 363–375 (2000). https://doi.org/10.1509/jmkr.37.3.363.18779
6. Arce-Cardenas, S., Fajardo-Delgado, D., Álvarez-Carmona, M.Á., Ramírez-Silva, J.P.: A tourist recommendation system: a study case in Mexico. In: Batyrshin, I., Gelbukh, A., Sidorov, G. (eds.) MICAI 2021. LNCS (LNAI), vol. 13068, pp. 184–195. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-89820-5_15
7. Bari, A., Brower, W., Davidson, C.: Using artificial intelligence to predict legislative votes in the united states congress. In: 2021 IEEE 6th International Conference on Big Data Analytics (ICBDA), pp. 56–60 (2021). https://doi.org/10.1109/ICBDA51983.2021.9403106
8. Blakeslee, S., et al.: Digital democracy project: making government more transparent one video at a time. Digit. Hum. (2015)
9. Budhwar, A., Kuboi, T., Dekhtyar, A., Khosmood, F.: Predicting the vote using legislative speech. In: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, pp. 1–10 (2018)
10. Burke, R.: Hybrid web recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web. LNCS, vol. 4321, pp. 377–408. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72079-9_12
11. California Legislative Information: AB-3235 public employees' retirement (2022). https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201720180AB3235
12. California Legislative Information: AB-976 electronic filing and service (2022). https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201720180AB976
13. California Legislative Information: California legislative information (2022). https://leginfo.legislature.ca.gov/
14. Cao, L.: Library personalized recommendation system based on collaborative filtering recommendation algorithm. In: Xu, Z., Alrabaee, S., Loyola-González, O., Zhang, X., Cahyani, N.D.W., Ab Rahman, N.H. (eds.) CSIA 2022. LNDECT, vol. 125, pp. 471–477. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-97874-7_61
15. Carrer-Neto, W., Hernández-Alcaraz, M.L., Valencia-García, R., García-Sánchez, F.: Social knowledge-based recommender system. application to the movies domain. Expert Syst. Appl. **39**(12), 10990–11000 (2012). https://doi.org/10.1016/j.eswa.2012.03.025. https://www.sciencedirect.com/science/article/pii/S0957417412004952
16. Chang, S.H., Abdul, A., Chen, J., Liao, H.Y.: A personalized music recommendation system using convolutional neural networks approach. In: 2018 IEEE International Conference on Applied System Invention (ICASI), pp. 47–49 (2018). https://doi.org/10.1109/ICASI.2018.8394293
17. Chen, A.: Context-aware collaborative filtering system: predicting the user's preference in the ubiquitous computing environment. In: Strang, T., Linnhoff-Popien, C. (eds.) LoCA 2005. LNCS, vol. 3479, pp. 244–253. Springer, Heidelberg (2005). https://doi.org/10.1007/11426646_23
18. De Meo, P., Quattrone, G., Ursino, D.: A decision support system for designing new services tailored to citizen profiles in a complex and distributed e-government scenario. Data Knowl. Eng. **67**(1), 161–184 (2008). https://doi.org/10.1016/j.datak.2008.06.005
19. Enda, J., Matsa, K.E., Boyles, J.L.: America's Shifting Statehouse Press: Can New Players Compensate for Lost Legacy Reporters? Pew Research Center (2014)

20. Esteban, B., Álvaro Tejeda-Lorente, Porcel, C., Arroyo, M., Herrera-Viedma, E.: Tplufib-web: a fuzzy linguistic web system to help in the treatment of low back pain problems. Knowl.-Based Syst. **67**, 429–438 (2014). https://doi.org/10.1016/j.knosys.2014.03.004. https://www.sciencedirect.com/science/article/pii/S0950705114000872
21. GeoPy Contributors: Geopy (2018). https://geopy.readthedocs.io/en/stable/
22. Gerrish, S., Blei, D.M.: The ideal point topic model: predicting legislative roll calls from text (2010)
23. Graefe, A., Bohlken, N.: Automated journalism: a meta-analysis of readers' perceptions of human-written in comparison to automated news. Media Commun. **8**(3), 50 (2020). https://doi.org/10.17645/mac.v8i3.3019
24. Gu, Y., Sun, Y., Jiang, N., Wang, B., Chen, T.: Topic-factorized ideal point estimation model for legislative voting network. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2014, pp. 183–192. Association for Computing Machinery, New York (2014). https://doi.org/10.1145/2623330.2623700
25. Howe, P., Robertson, C., Grace, L., Khosmood, F.: Exploring reporter-desired features for an AI-generated legislative news tip sheet. ISOJ **12**(1), 17–44 (2022)
26. Iftikhar, A., Ghazanfar, M.A., Ayub, M., Mehmood, Z., Maqsood, M.: An improved product recommendation method for collaborative filtering. IEEE Access **8**, 123841–123857 (2020). https://doi.org/10.1109/ACCESS.2020.3005953
27. Klimashevskaia, A., Gadgil, R., Gerrity, T., Khosmood, F., Gütl, C., Howe, P.: Automatic news article generation from legislative proceedings: a phenom-based approach. In: Espinosa-Anke, L., Martín-Vide, C., Spasić, I. (eds.) SLSP 2021. LNCS (LNAI), vol. 13062, pp. 15–26. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-89579-2_2
28. Kumar BP, V., VS, P., et al.: Survey on classification and summarization of documents. In: Proceedings of the Second International Conference on Emerging Trends in Science & Technologies For Engineering Systems (ICETSE-2019), pp. 7–13 (2019)
29. Lee, J., Kothari, N., Natsev, P.: Content-based related video recommendations. In: Advances in Neural Information Processing Systems (NIPS) Demonstration Track (2016). http://www.joonseok.net/papers/video_recs_demo.pdf
30. Lu, J., Wu, D., Mao, M., Wang, W., Zhang, G.: Recommender system application developments: a survey. Decis. Support Syst. **74**, 12–32 (2015)
31. Miller, G.A.: Wordnet: a lexical database for english. Commun. ACM **38**(11), 39–41 (1995). https://doi.org/10.1145/219717.219748
32. Moens, M., Logghe, M., Dumortier, J.: Legislative databases: current problems and possible solutions. Int. J. Law Inf. Technol. **10**(1), 1–22 (2002). https://doi.org/10.1093/ijlit/10.1.1
33. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72079-9_10
34. Prolific: Prolific (2022). https://prolific.co
35. Rani, S., Singh, J.: Enhancing Levenshtein's edit distance algorithm for evaluating document similarity. In: Sharma, R., Mantri, A., Dua, S. (eds.) ICAN 2017. CCIS, vol. 805, pp. 72–80. Springer, Singapore (2018). https://doi.org/10.1007/978-981-13-0755-3_6

36. Ruprechter, T., Khosmood, F., Guetl, C.: Deconstructing human-assisted video transcription and annotation for legislative proceedings. Digit. Gov. Res. Pract. **1**(3) (2020). https://doi.org/10.1145/3395316
37. Salloum, S.A., Khan, R., Shaalan, K.: A survey of semantic analysis approaches. In: Hassanien, A.-E., Azar, A.T., Gaber, T., Oliva, D., Tolba, F.M. (eds.) AICV 2020. AISC, vol. 1153, pp. 61–70. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-44289-7_6
38. Shani, G., Gunawardana, A.: Evaluating recommendation systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 257–297. Springer, Boston, MA (2011). https://doi.org/10.1007/978-0-387-85820-3_8
39. Shaw, A.: As statehouse press corps dwindles, other reliable news sources needed. Better Government Association (2017)
40. Stanchev, L.: Creating a similarity graph from wordnet. In: Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14), WIMS 2014. Association for Computing Machinery, New York (2014). https://doi.org/10.1145/2611040.2611055
41. Tarus, J.K., Niu, Z., Mustafa, G.: Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. Artif. Intell. Rev. **50**(1), 21–48 (2018)
42. Terán, L., Meier, A.: A fuzzy recommender system for eelections. In: Andersen, K.N., Francesconi, E., Grönlund, Å., van Engers, T.M. (eds.) EGOVIS 2010. LNCS, vol. 6267, pp. 62–76. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15172-9_6
43. Thomas, M., et al.: Quick check: a legal research recommendation system. In: NLLP@ KDD, pp. 57–60 (2020)
44. Winkler, M.A.: California poised to overtake Germany as world's no. 4 economy. Bloomberg (2022)
45. Yano, T., Smith, N.A., Wilkerson, J.D.: Textual predictors of bill survival in congressional committees. In: Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Montréal, Canada, pp. 793–802. Association for Computational Linguistics (2012). https://aclanthology.org/N12-1097
46. Yapriady, B., Uitdenbogerd, A.L.: Combining demographic data with collaborative filtering for automatic music recommendation. In: Khosla, R., Howlett, R.J., Jain, L.C. (eds.) KES 2005. LNCS (LNAI), vol. 3684, pp. 201–207. Springer, Heidelberg (2005). https://doi.org/10.1007/11554028_29