Edit Categories and Editor Role Identification in Wikipedia

Diyi Yang♣, Aaron Halfaker◇, Robert Kraut♠, Eduard Hovy♣

♣ Language Technologies Institute, Carnegie Mellon University {diyi,hovy}@cmu.edu
◇ Wikimedia Foundation ahalfaker@wikimedia.org
♠ Human-Computer Interaction Institute, Carnegie Mellon University robert.kraut@cmu.edu

Abstract

In this work, we introduced a corpus for categorizing edit types in Wikipedia. This fine-grained taxonomy of edit types enables us to differentiate editing actions and find editor roles in Wikipedia based on their low-level edit types. To do this, we first created an annotated corpus based on 1,996 edits obtained from 953 article revisions and built machine-learning models to automatically identify the edit categories associated with edits. Building on this automated measurement of edit types, we then applied a graphical model analogous to Latent Dirichlet Allocation to uncover the latent roles in editors' edit histories. Applying this technique revealed eight different roles editors play, such as Social Networker, Substantive Expert, etc.

Keywords: Edit Category, Role Identification, Topic Modeling

1. Introduction

Distributed work teams in online communities have become increasingly important in creating innovative products, such as GNU, Linux and Wikipedia. Millions of volunteers participate in the online production communities, exchange their expertise and ideas, and collaborate to produce complex artifacts. For example, in Wikipedia, editors take up different responsibilities, when editing articles, based on their interest and expertise. Some, for example, might add substantive new content to articles while others may focus on copy-editing. Better understanding of the participants and what they do, how they behave can make these online communities more successful. Thus, the problem is, given two revisions of a piece of text, what actions have been done by a user to transform the original version into the new one, and who that user is? Our goal in this work is to develop fine-grained taxonomy to categorize editors' elemental actions when editing Wikipedia articles and design new methods to identify roles that editors exhibit.

The problem of identifying editors' roles in Wikipedia has attracted significant attention (Arazy et al., 2015; Ferschke et al., 2015). Numerous studies have discussed how to identify roles based on users' behavioral regularities and social network signatures (Welser et al., 2011; Welser et al., 2007). Most research classifies editors based either on their edits in different namespaces or via the user attributes such as access privileges (Arazy et al., 2015), barnstars (Kriplean et al., 2008), etc. Classification based on users' attributes is relatively accurate, but this information is not available for many active editors and is insufficient in explaining the nature of an editor's work. While classification based on edit histories can be constructed for most active editors, current approaches focus on simple edit counts and access privileges fail to provide a finer grained description of the work actually performed in an edit. For example, it cannot tell the difference between an editor who copy-edits a paragraph and an editor who inserts markup or template to an article.

In this work, we extend Daxenberger's fine grained taxonomy of edit types (Daxenberger and Gurevych, 2012; Daxenberger and Gurevych, 2013) to differentiate edits and editors who occupy different editing roles. The edits are distinguished contextually in terms of the object being edited (e.g. information, template, reference, etc.) and functionally, in terms of the edit operation (e.g. insert, delete, etc.). The corpus construction will be described in detail later. Based on our corpus, we then described the development of methods for the automated measurement of these 24 edits categories revealed in users' edits. These categories can be identified with a relatively reasonable performance by using a multi-label classification algorithm.

Building on this automated measurement of edit types, we use a graphical model analogous to LDA topic modeling to identify the latent roles editors occupy, much as document comprise topics. In contrast to studies that employed either clustering analysis or principle component analysis to extract user roles (Liu and Ram, 2011), our role modeling treats an editor as comprising multiple roles at the same time. This approach makes the role more interpretable in capturing the versatility and dynamics of editors. This approach identified eight editor roles in Wikipedia: Social Networker, Fact Checker, Substantive Expert, Wiki Gnome, Vandal Fighter, Fact Updater and Wikipedian.

2. Corpus Construction

Basing our research on Daxenberger et al. (Daxenberger and Gurevych, 2012; Daxenberger and Gurevych, 2013), we distinguished between revisions and edits. Here, a *revision* is created whenever an editor makes changes to a Wikipedia page. An *edit* is a coherent local change and regarded as one single editing action. Each edit is associated with a set of labeling of edit categories, representing in which aspects it has been changed. A revision can contain a set of distinct local edits. For each pair of adjacent revisions, we collected a set of edits that has been made to transform from its parent revision into this revision.

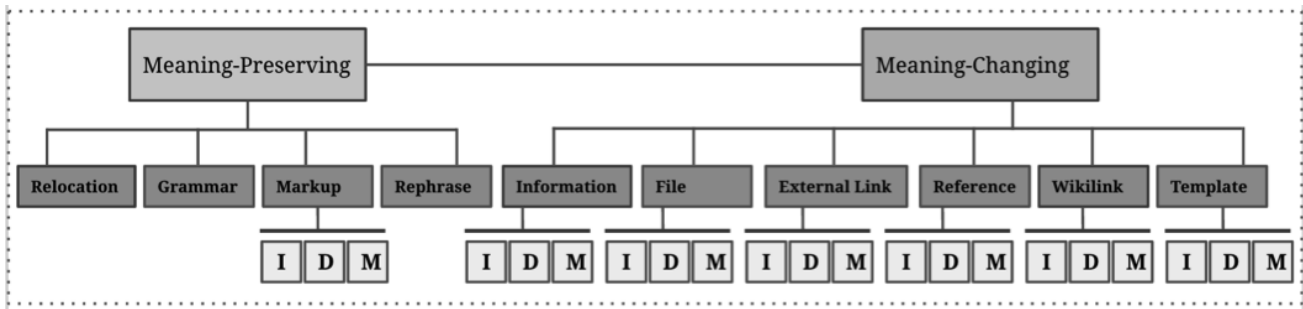Figure 1 provides an overview of our edit taxonomy. In this work, we annotated a set of edits rather than revisions.

Figure 1: **The Taxonomy of Edit Categories**. Note: Insertion is abbreviated as I, Deletion as D and Modification as M

In general, this taxonomy considers actions (insert, delete, modify) applied to different objects in Wikipedia (e.g., information, templates or references), leading to 24 distinct edit types. The two top-level layers summarize whether these edit categories are meaning-preserving or meaning-changing.

Of the meaning-preserving edits, Grammar (G) means the edit is correcting spelling or grammatical errors, as well as fixing punctuation. When an edit attempts to paraphrase words or sentences, it is categorized as Rephrase (P); if such edit only moves entire lines without any other changes, it is considered as Relocation (R). For edits that try to operate with the markup segments, such as ''''' or HTML tags, depending how it affects the markup, we divide them into three sub-categories: Markup Insertion (M-I), Markup Deletion (M-D), and Markup Modification (M-M).

Similarly, depending on the context being modified and how an edit affects the context, we divided Meaning-Changing edits into 18 categories, with three basic operations (Insertion, Deletion, Modification) associated with six context (Information, File, Template, External link, Reference, Wikilink). In detail, Information refers to whether an edit affects the information content or sentence meaning; File points to file objects which are usually in the form of a `[[File:...]]` link, but sometimes also appear in a template as a parameter `filename=File:....`. Template refers to template transclusions, e.g. `{{citationneeded}}` or `{{Infobox|...}}`. External Link means links that use absolute URLs to point towards external domains, such as "[https://wikimedia.org The Wikimedia Portal]". Reference stands for a reference tag and/or a citation template, such as `<ref>{{citebook|...}}</ref>`. Wikilinks are links to another page within English Wikipedia, e.g. `[[John_Smith_(professor)|Dr.Smith]]`. It is worth mentioning that, we break Daxenberger's 'Reference' category (Daxenberger and Gurevych, 2012; Daxenberger and Gurevych, 2013) into three finer-grained categories: External Link [1] refers to links from articles to web pages outside Wikipedia; Wikilink [2] refers to links to another page within the English Wikipedia and Reference describes the source of the information, to help the reader who wishes to verify it, or to pursue it in greater depth[3]. This results in: Information Insertion (I-I), Information Deletion (ID), and Information Modification (I-M), Template Insertion (T-I), Template Deletion (T-D), Template Modification (T-M), File Insertion (F-I), File Deletion (F-D), File Modification (F-M), External Link Insertion (E-I), External Link Deletion (E-D), External Link Modification (E-M), Reference Insertion (R-I), Reference Deletion (R-D), Reference Modification (R-M), Wikilink Insertion (W-I), Wikilink Deletion (W-D), and Wikilink Modification (W-M).

The dataset can be downloaded here[4]. This dataset contains 1,996 edits, randomly sampled from sampled 953 revisions from June 10th, 2014 to June 10th, 2015. It contains 729 distinct editors in total. We eliminated revisions made by editors who had fewer than 15 edits. Anonymous users are included. We annotated the corpus based on a written annotation guideline. The annotation task is framed as a multi-label classification. That is, each edit will be assigned to one or more edit categories. For example, if an edit added a sentence to an article, this edit might involve insertion of information only or the insertion of information, a Wikilink insertion and a reference simultaneously. A revision edit containing the three components would be multi-labeled as I-I, W-I and R-I.

To assess the validity of the annotation, we compared the annotations of 63 randomly sampled revision edits made by the first author and by an expert Wikipedian. Despite the difference in Wikipedia editing experience between the hand coders, the agreement between the annotations was substantial (Cohen's Kappa = 0.723; see (Landis and Koch, 1977)).

3. Applications of Edit Categories

With the recent popularity of online collaboration platforms such as Wikipedia, figuring out what has been done to transform one version of a text to another version becomes more and more important. Such editing process reflects both the intentions behind a textual change and interaction and collaboration between users. To assist our understanding towards this writing process and facilitate further applications building upon the revision data, we introduced the above

---

Edit Category corpus, a fine grained taxonomy which contains 24 categories of elemental actions in the article editing processes of Wikipedia. This taxonomy of edit categories and the constructed corpus can assist several important tasks and applications:

1. **Editor Role Identification**: A deeper analysis to the type of work done by different editors might reflect who they are. For example, a user who always works on copy editing and rephrasing might be a copy-editor; one seems to be a substantive expert if he/she contributes to information insertion a lot. Therefore, our edit taxonomy enables us to identity user roles based on their elemental actions.

2. **Collaboration Quality Prediction**: The quality of collaboration varies widely. For example, although there are over 4.5 million articles in the English Wikipedia, as of September 2014, Wikipedians have evaluated fewer than 0.1% of them as good articles or better and over 88% of them as start or stub[5] class articles. Collaboration among editors with different skills is essential to developing high quality articles(Kittur and Kraut, 2008). Thus, determining how contribution by different types of work and by different editors at distinct times in an article's history influence changes in its quality is of great use to better understand the causes of quality variance in Wikipedia (De la Calzada and Dekhtyar, 2010).

3. **Quality Flaw Detection**: The detection and improvement of low quality information is an essential component in online production communities. Different from existing studies in quality flaw prediction (Anderka et al., 2012; Anderka et al., 2011; Ferschke et al., 2013), our taxonomy enables us to provide which specific aspects (information, reference, wikilink, template or markup, etc.) of an article needs improvement and what operations should be performed.

4. **Domain Adaption**: Even though the edit taxonomy introduced above is for English Wikipedia, it can be applied to other language versions of Wikipedia. For example, similar taxonomy, or even automated classification models can be transformed for another language. Beyond the context of Wikipedia, similar taxonomies can be designed for analyzing the collaboration and interaction happened in other online contexts such as Google Docs, ShareLatex[6] or Github[7], etc.

4. Example: Identification of Edit Categories and Roles

4.1. Edit Categories Prediction

In the above section, we listed several important tasks, which can be built on our edit taxonomy and the corresponding annotated corpus. In this part, as an example of

utilizing the corpus, we designed a machine-learning model to automatically identify the edit categories associated with edits. Specifically, the goal was to classify an edit into one or more of the edit categories based on characteristics of the text changed, the comments editors used to describe their edits, and characteristics of the edit. To capture these characteristics, we developed a set of features, including whether a revision is marked as minor change, comment length, whether the author is registered or IP user, number of edits in this revision, where (reference, template, file, external link, wikilink, markup) a revision is performed, etc. Given the input feature representation of an edit, we then built a machine-learning model for this multi-label classification. We utilized the Revision Scoring package[8] to collect Relocation edits, and did not include the category of relocation in our prediction stage.

In detail, we used two of the multi-label classifier implemented in Mulan (Tsoumakas et al., 2011) with ten cross validation. We used the RAkEL ensemble method classifier, described in (Su and Rousu, 2015). This method randomly chooses a small subset with k categories from the overall set of categories. We compared this with the MLkNN classifier, which is based on K Nearest Neighbor method. Table 1 shows the evaluation metrics including Recall, Precision, micro-averaged F1 score and AUC (Area under Curve). Both methods gave classifications that agreed with the human judgments, indicated by the AUC score of 0.865 and 0.906 respectively. We chose to use RAkEL method in order to acquire a relatively better performance in terms of F1 score. Figure 2 describes the performance of classification on different edit categories.

	Recall	Precision	F1	AUC
RAkEL	0.575	0.730	**0.643**	0.865
MLkNN	0.363	0.724	0.482	0.906

Table 1: Edit Categories Prediction Results

4.2. Editor Role Identification

As an example, we performed the task of editor role identification based on the trained edit categories classifier on this corpus. Besides editors' edit types performed in articles, we also included the number of edits editors' made in each Wikipedia namespace[9] into the role models. We also include the number of reverts (i.e., returning a Wikipedia to a prior state) and vandalistic edits editors made in the role model. We take advantage of two utilizes written by the Wikimedia Foundation that accurately measure this activity. Mediawiki-utilities Revert Check API[10] measures revert. The Vandalism API[11] returns the probability that a given revision is vandalism.

---

[5]https://en.wikipedia.org/wiki/Wikipedia:Stub

[6]https://www.sharelatex.com

[7]https://github.com/

[8]http://pythonhosted.org/revscoring/index.html

[9]https://en.wikipedia.org/wiki/Wikipedia:Namespace

[10]https://pythonhosted.org/mediawiki-utilities/lib/reverts.html#mw.lib.reverts.api.check

[11]http://ores.wmflabs.org/scores/enwiki/?models=reverted&revids=revision_id
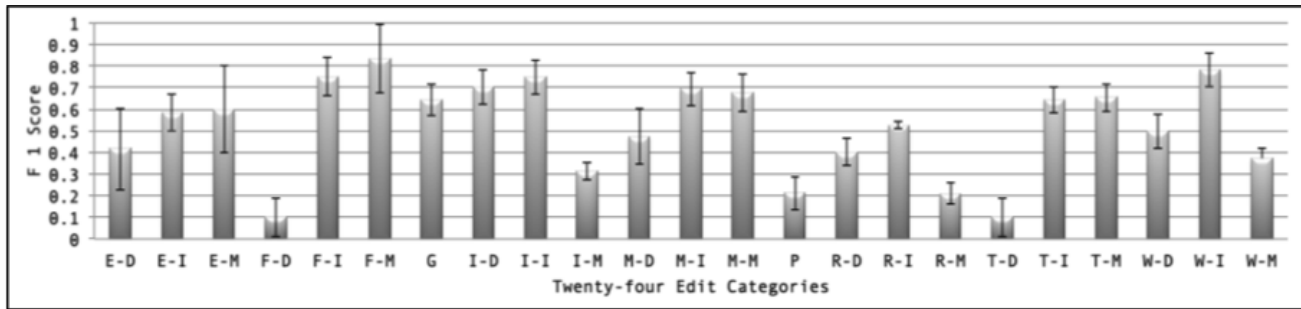
Figure 2: **F1 Score Performance over Different Edit Categories**

Derived Roles	Representative Behavior
Social Networker	Main talk namespace, user namespace, reference modification
Fact Checker	Information deletion, wikilink deletion, reference deletion, file deletion, markup deletion external link deletion
Substantive Expert	Information insertion, wikilink insertion, markup insertion, reference insertion, external link insertion, file insertion, template insertion
Copy Editor	Grammar, paraphrase, relocation
Wiki Gnomes	Wikilink modification, Template insertion, markup modification, Wikipedia talk namespace, category namesapce
Vandal Fighter	Reverting, user talk namespace, reference insertion, external link deletion, paraphrase
Fact Updater	Template modification, reference modification, file namespace
Wikipedian	Wikilink insertion, Wikipedia namespace, template namespace, file insertion

Table 2: Derived Editor Roles and Their Representative Edit Types

Our objective is to identity the roles that editors play, clustering editors who share patterns of work, using the types of edit they make in articles, their revert and vandalism, and edit counts in other namespaces. For this purpose, we used the graphic model underlying the Latent Dirichlet Allocation (LDA) method (Blei et al., 2003), where edit types, reverts, vandalistic edit and edits in non-article namespaces are regarded as 'word'; an editor's edit history of 'word' is analogous to 'document'. The latent roles derived are analogous to an LDA topic.

We trained a LDA model on a dataset that consists of 38,520 editors and 626,761 revisions. The revision data is randomly sampled during Dec 1st, 2014 to Jan 1st, 2015 time period. We experimented with driving from 5 to 15 roles (i.e., topics in the LDA software) and in the end decided to derive eight latent roles because they were more interpretable than other solutions. We selected the edit-types and namespaces that are most likely to correspond to that topic and visualized the sampled edit categories, revert and vandalism, and edits to other namespaces for each derived role. Two experts familiar with Wikipedia applied a label to each topic, based on the behaviors most heavily associated with each role, and came up with eight roles as shown in Table 2.

5. Discussion and Future Work

5.1. Discussion

This paper described a corpus for categorizing edit types in Wikipedia and several important tasks of utilizing the annotated corpus. In detail, we introduced a fine-grained taxonomy of edit types to characterize users' edits and built machine learning models to automatically identify the edit

categories in each edit. We appropriated LDA-like graphical models to extracted latent roles from editors' history of edit activities. We made this corpus public and hope it will assist and encourage work on automatic identification of elemental actions and social roles.

5.2. Future Work

We would like to investigate the following two directions as future work:

1. **Taxonomy Expansion**: The current edit taxonomy is incapable of capturing the intentions of users' elemental editing actions. For example, 'information insertion' should be differentiated between adding substantive new content and explaining an existing fact; and information modification might be either copy editing or removing bias for neutral point of view. To incorporate more semantic information, we are developing a two layer taxonomy[12] that considers both semantic intentions and syntactic operations. It distinguishes edits semantically in terms of the intention of editing (e.g. elaboration, clarification, verifiability, etc.), contextually in terms of the object being edited (e.g. information, template, reference, etc.), and functionally, in terms of the edit operation (e.g. insert, delete, etc.). Such expanded taxonomy will provide us with better opportunities for understanding user behaviors such as what they did and why they do them.

2. **Unified Taxonomy**: Similar taxonomies might be designed for analyzing the collaboration processes in

---

[12] https://en.wikipedia.org/wiki/Wikipedia:
Labels/Edit_types/Taxonomy

other online contexts such as Google Docs, ShareLatex or Github, etc. Instead of designing different taxonomies for specific context, we plan to build a unified framework to automatically learn the latent representation of the taxonomy and its elemental actions.

6. Acknowledgements

7. Bibliographical References

Anderka, M., Stein, B., and Lipka, N. (2011). Detection of text quality flaws as a one-class classification problem. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 2313–2316. ACM.

Anderka, M., Stein, B., and Lipka, N. (2012). Predicting quality flaws in user-generated content: the case of wikipedia. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 981–990. ACM.

Arazy, O., Ortega, F., Nov, O., Yeo, L., and Balila, A. (2015). Functional roles and career paths in wikipedia. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 1092–1105, New York, NY, USA. ACM.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Daxenberger, J. and Gurevych, I. (2012). A corpus-based study of edit categories in featured and non-featured Wikipedia articles. In *Proceedings of COLING 2012*, pages 711–726, Mumbai, India, December. The COLING 2012 Organizing Committee.

Daxenberger, J. and Gurevych, I. (2013). Automatically classifying edit categories in Wikipedia revisions. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 578–589, Seattle, Washington, USA, October. Association for Computational Linguistics.

De la Calzada, G. and Dekhtyar, A. (2010). On measuring the quality of wikipedia articles. In *Proceedings of the 4th Workshop on Information Credibility*, WICOW '10, pages 11–18, New York, NY, USA. ACM.

Ferschke, O., Gurevych, I., and Rittberger, M. (2013). The impact of topic bias on quality flaw prediction in wikipedia. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 721–730, Sofia, Bulgaria, August. Association for Computational Linguistics.

Ferschke, O., Yang, D., and Rose, C. (2015). A lightly supervised approach to role identification in wikipedia talk page discussions.

Kittur, A. and Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: Quality through coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, pages 37–46, New York, NY, USA. ACM.

Kriplean, T., Beschastnikh, I., and McDonald, D. W. (2008). Articulations of wikiwork: Uncovering valued work in wikipedia through barnstars. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, pages 47–56, New York, NY, USA. ACM.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.

Liu, J. and Ram, S. (2011). Who does what: Collaboration patterns in the wikipedia and their impact on article quality. *ACM Trans. Manage. Inf. Syst.*, 2(2):11:1–11:23, July.

Su, H. and Rousu, J. (2015). Multilabel classification through random graph ensembles. *Machine Learning*, 99(2):231–256.

Tsoumakas, G., Spyromitros-Xioufis, E., Vilcek, J., and Vlahavas, I. (2011). Mulan: A java library for multilabel learning. *The Journal of Machine Learning Research*, 12:2411–2414.

Welser, H. T., Gleave, E., Fisher, D., and Smith, M. (2007). Visualizing the Signatures of Social Roles in Online Discussion Groups. *The Journal of Social Structure*, 8(2).

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., and Smith, M. (2011). Finding social roles in wikipedia. In *Proceedings of the 2011 iConference*, iConference '11, pages 122–129, New York, NY, USA. ACM.