GraM-SUM: A Graph-Based Model for Book-Level Summarization

Romain L. Hardy
UC Berkeley School of Information
rhardy@berkeley.edu

Abstract

Although summarization is a well-studied problem, most related work focuses on short texts: newspaper articles, academic papers, and short stories. By contrast, long-text summarization remains mostly unexplored. In this paper we introduce GraM-SUM, a graph-based model that produces summaries of book-length texts by combining chains of low-level summaries in a way that considers progression, diversity, and importance. We show that GraM-SUM generally outperforms a perplexity-based approach at the chapter and book levels on the BOOKSUM dataset, and produces summaries with more diverse sentence-level embeddings.

1 Introduction

Summarization is a long-standing task in the field of natural language processing. Early approaches like TextRank (Mihalcea, 2004) focused on extractive methods for selecting representative sentences within a body of text. In 2017, the introduction of transformers (Vaswani et al., 2017) and "BERTology" spawned a large family pre-trained language models that could be fine-tuned to accomplish a wide variety of specialized tasks, including summarization. While a substantial body of work exists concerning the summarization of newspaper articles, academic papers, and short texts, comparatively few studies address book-length summarization. This is in part due to the unique challenges that this level of summarization presents. For a book-length summarization model to be effective, it must be able to capture long-range textual and semantic relationships, sometimes across thousands of tokens. In theory, a transformer model could capture such relationships by attending to tokens across long spans. However, since the memory consumption of transformers scales quadratically with the number of input tokens, they are limited in their ability to process large texts. Therefore, book-length summarization models must be creative in the ways that they process text to generate new summaries.

In this work we introduce GraM-SUM, a graph-based model for summarizing book-length texts. This model operates at three separate levels of hierarchy: paragraphs, chapters, and books. At the paragraph level, the model uses a `t5-small` encoder-decoder to generate summaries. At the chapter and book levels, we adapt a graph-based algorithm (Gorinski and Lapata, 2015) that selects chains of low-level summaries with maximal progression, diversity, and importance. We show that when the chain length is small, this approach outperforms and produces more diverse sentence embeddings than a perplexity-based approach.

2 Datasets

2.1 Summarization

In order to train and evaluate our summarization model, we leverage the recently published BOOKSUM dataset (Kryściński et al., 2021), a collection of publicly available novels, plays, and stories from *Project Gutenberg*. Reference summaries are scraped from various independent sources such as *Shmoop*, *SparkNotes*, and *GradeSaver*. The authors align texts and summaries at the paragraph, chapter, and book levels. In total, the dataset contains 142,753 paragraph-level, 12,293 chapter-level, and 436 book-level text-summary pairs.

2.2 Named entity recognition

Part of GraM-SUM involves resolving entities in texts in order to form graphs. For this task, we fine-tune a named entity recognition model (NER) on the CoNLL 2003 dataset (Tjong Kim Sang and De Meulder, 2003). This dataset consists of 20,744 passages and their associated entity tags: PER for person, LOC for location, ORG for organization, and MISC for miscellaneous.

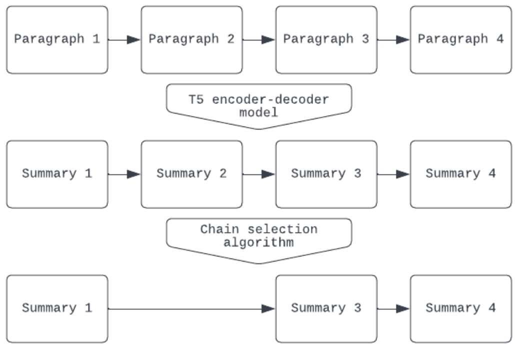

Figure 1: A visualization of the chapter-level logic used by GraM-SUM. First, the paragraph-level encoder-decoder summarizes each paragraph in the chapter. Then, a chain selection algorithm selects an optimal chain of paragraph summaries to form the chapter summary.

3 Past Work

Several attempts, both extractive and abstractive, have been made in regards to long-form summarization. An early attempt came in 2004 with the introduction of TextRank (Mihalcea, 2004), an extractive model inspired by the PageRank algorithm. In 2015, Gorinski and Lapata (2015) developed an algorithm for selecting an optimal chain of scenes from a movie script. Later, new attention-based architectures were proposed for processing long texts, such as hierarchical attention networks (Yang et al., 2016) and the "Longformer" (Beltagy et al., 2020). Recently, OpenAI presented a model that recursively summarizes book-length texts, first by summarizing small chunks, then concatenating the summaries, summarizing those, and so on (Wu et al., 2021).

4 Methods

4.1 Paragraphs

At the paragraph level, we fine-tune a `t5-small` encoder-decoder model for 3 epochs on the BOOKSUM training dataset. When generating summaries for new paragraphs, we use beam search with a beam size of 3, a minimum length of 30, and a maximum length of 200.

4.2 Chapters

To produce chapter-level summaries, we adapt an algorithm created by Gorinski and Lapata (2015) to summarize movie scripts. At a high level, the algorithm selects an optimal chain $C_{\text{opt}} = \{s_1, \ldots, s_n\}$ of paragraph-level summaries s_i and aggregates it into a chapter-level summary, as illustrated in Figure 1.

In their work, Kryściński et al. (2021) form chains consisting of the top-*k* paragraph summaries ranked by perplexity. Although this approach incentivizes coherent summaries from a language modeling perspective, it sometimes ignores candidate summaries that, in spite of their higher perplexities, contribute new and meaningful information to the summary. Instead, our algorithm selects the chain with maximal progression P, diversity D, and importance I:

$$C_{\text{opt}} = \arg\max_{C} Q(C), \tag{1}$$

$$Q(C) = \lambda_P P(C) + \lambda_D D(C) + \lambda_I I(C), \tag{2}$$

where λ_P, λ_D, and λ_I are hyperparameters. We define these terms in the following sections.

Progression We operationalize progression from a graph-theoretic perspective. A chapter is represented by a bipartite graph $B = (V, E)$. The set of vertices $V = S \cup \mathcal{E}$ is the union of the paragraph summaries S for that chapter and the named entities $\mathcal{E}$ identified in S. The edges $E = \{w_{e\to s}\} \cup \{w_{s\to e}\}$ are directed weights between each entity-summary pair. In total, B contains one vertex per summary, one vertex per entity, and two directed edges per entity-summary pair. An example of a bipartite graph is shown in Figure 2.

Entities to place in B are identified using a BERT NER model fine-tuned on the CoNLL 2003 dataset. For each text in the BOOKSUM evaluation set, we extract the 100 most frequent entities belonging to the PER, LOC, and ORG tags and having classification scores greater than 0.99. Because some identified entities may be duplicated (i.e. "Harry" and "Harry Potter"), we adopt a deduplication strategy that matches each identified entity with the closest other entity by token set ratio. If the ratio between two entities exceeds 70, then we link those entities together; otherwise, we keep them separate.

The weight $w_{e\to s}$ represents the global importance of entity e to summary s, relative to other summaries:

$$w_{e\to s} = \begin{cases} 1/N_{e\to s} & e \in s \\ 0 & e \notin s, \end{cases} \tag{3}$$

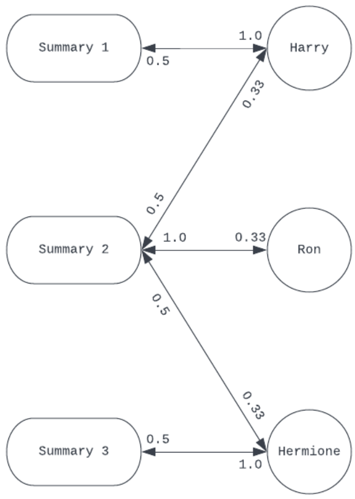

Figure 2: An example of a bipartite graph. Each entity node (right) is connected to one or multiple summaries (left). The weights of each outgoing edge are normalized so that they sum to 1.

where $N_{e \to s}$ is the number of summaries in which e appears. Effectively, the importance of an entity is spread equally across all summaries in which it is mentioned. On the other hand, $w_{s \to e}$ represents the importance of entity e within s relative to other entities:

$$w_{s \to e} = \begin{cases} N_{s \to e} / \sum_{e' \in p} N_{s \to e'} & e \in s \\ 0 & e \notin s, \end{cases} \tag{4}$$

where $N_{s \to e}$ is the number of times entity e appears in summary s. The more often an entity appears in a summary relative to other entities, the higher the value of $w_{s \to e}$.

Intuitively, the progression of a chain C of paragraph-level summaries is a measure of the narrative coherence of that chain. If a paragraph-level summary in the chain includes an entity, then subsequent summaries in the chain should include related entities. Formally, we say that a tuple of paragraph-level summaries (s_i, s_{i+1}) has a high degree of progression if a random walker starting at paragraph summary s_i in B has a high probability of ending at s_{i+1}. Let r_{s_i} be the stationary distribution of a random walker with restart starting at summary s_i. r_{s_i} satisfies

$$r_{s_i} = \epsilon (I - (1 - \epsilon)T)^{-1} a_{s_i}, \tag{5}$$

where T is the transition matrix for B, ϵ is the probability that the random walker restarts, and a_{s_i} is a vector whose s_i-th entry is 1 and the rest are zero. Notably, the entry $r_{s_i}[s_{i+1}]$ represents the probability that the random walker starts at s_i and ends at s_{i+1}.

Now let $r^e_{s_i}$ be the stationary distribution for the same random walker, but with outgoing connections from entity e removed. Here, $r^e_{s_i}[s_{i+1}]$ represents the probability that the random walker ends at node s_{i+1}, but without the influence of entity e. Naturally, we can define the influence of e on the progression from s_i to s_{i+1} as the difference between $r_{s_i}[s_{i+1}]$ and $r^e_{s_i}[s_{i+1}]$. Summing over all entities and all neighboring pairs of paragraph summaries in C, we arrive at the full progression score for C:

$$P(C) = \sum_{i=1}^{n-1} \sum_{e \in \mathcal{E}_i} r_{s_i}[s_{i+1}] - r^e_{s_i}[s_{i+1}].$$

Diversity Diversity encourages chains of paragraph summaries that are dissimilar from one another, so that the same information is not repeated. It is composed of two terms, an entity overlap term d_{over} and a sentiment overlap term d_{sent}. d_{over} measures the degree to which entities are shared between neighboring pairs of summaries:

$$d_{\text{over}}(s_i, s_{i+1}) = 1 - \left| \frac{\mathcal{E}_i \cap \mathcal{E}_{i+1}}{\mathcal{E}_i \cup \mathcal{E}_{i+1}} \right|. \tag{6}$$

The more entities are shared between the summaries, the greater their similarity and the lower the value of d_{over}. Meanwhile, d_{sent} measures the sentiment overlap between neighbors in the chain. Gorinski and Lapata (2015) use a pre-trained lexicon to predict sentiments, but we use a sentence transformer W instead. We determine the average sentence embedding for each paragraph summary, then compute the dot product similarity of those embeddings for neighboring pairs:

$$d_{\text{sent}}(s_i, s_{i+1}) = W(s_i)^\top W(s_{i+1}). \tag{7}$$

The full diversity score for chain C is the average of the entity overlap and sentiment scores across all neighboring pairs:

$$D(C) = \frac{1}{2} \sum_{i=1}^{n-1} d_{\text{over}}(s_i, s_{i+1}) + d_{\text{sent}}(s_i, s_{i+1}). \tag{8}$$

Importance We consider a summary s_i to be important if it contains globally important entities. As a proxy, we assume an entity is important if it occurs frequently throughout the corresponding book-level text. Therefore, a natural way to define the importance of chain C is as the sum of the book-level frequencies of all unique entities $\mathcal{E}_C$ in the summaries composing C:

$$I(C) = \frac{1}{\sum_{e \in \mathcal{E}} N_e} \sum_{e \in \mathcal{E}_C} N_e, \tag{9}$$

where N_e is the number of times entity e appears in the full text. The maximum importance score is 1, which occurs if C contains all entities found in the full text.

Chain Selection For a fixed chain length k, GraM-SUM aims to solve Equation 1. When the number of paragraph-level summaries is small, we can search the space of possible chains exhaustively for C_{opt}. However, the size of the search space grows combinatorially with the number of paragraph-level summaries. For example, a chapter composed of 50 paragraphs has 2,118,760 possible chains of length 5. To render the run-time of the model more reasonable, we only consider a random sample of 10,000 chains.

4.3 Books

Our methodology for producing book-level summaries is identical in principle to that for producing chapter-level summaries. However, instead of constructing chains of paragraph-level summaries to create chapter-level summaries, we construct chains of chapter-level summaries to create book-level summaries.

5 Results

	Perplexity (Chapter)			**GraM-SUM (Chapter)**		
k	R-1$_{F_1}$	R-2$_{F_1}$	R-L$_{F_1}$	R-1$_{F_1}$	R-2$_{F_1}$	R-L$_{F_1}$
2	20.79	3.85	12.15	22.67	4.13	12.81
3	26.34	4.98	14.24	27.84	5.21	14.73
4	30.14	5.80	15.60	31.38	6.04	15.97
5	32.68	6.41	16.43	33.69	6.59	16.69
6	34.48	6.85	16.94	35.34	7.01	17.15
7	35.89	7.21	17.35	36.58	7.34	17.51
8	36.89	7.46	17.63	37.43	7.58	17.75
9	37.71	7.72	17.89	38.14	7.81	17.98
10	38.33	7.91	18.06	38.53	7.92	18.05
	Perplexity (Book)			**GraM-SUM (Book)**		
k	R-1$_{F_1}$	R-2$_{F_1}$	R-L$_{F_1}$	R-1$_{F_1}$	R-2$_{F_1}$	R-L$_{F_1}$
2	23.50	4.73	11.47	29.33	5.51	12.80
3	29.84	6.09	13.46	35.74	6.87	14.39
4	34.63	7.09	14.57	38.73	7.70	15.00
5	37.37	7.60	15.02	40.19	8.24	15.19
6	39.37	8.02	15.50	41.24	8.58	15.37
7	40.32	8.29	15.55	41.37	8.74	15.27
8	40.79	8.59	15.55	41.31	8.80	15.10
9	40.82	8.73	15.59	40.78	8.85	14.97
10	40.50	8.40	15.70	40.57	8.80	15.45

Table 1: A comparison of GraM-SUM against the perplexity-based approach at the chapter (top) and book (bottom) levels on the BookSum dataset.[1]

Table 1 compares GraM-SUM against the perplexity-based baseline at the chapter and book levels on the BookSum dataset. We report ROUGE-n F_1 scores, which are computed with stemming.

In general, we notice that the ROUGE scores for both GraM-SUM and the perplexity-based model tend to increase with chain length, although with diminishing returns. This signifies that while chaining low-level summaries is useful for producing high-level summaries, some of the information becomes redundant in long chains.

Interestingly, we do not observe significant changes in performance as we vary the weights of the progression, diversity, and importance scores in Equation 1. By comparison, the choice of beam search hyperparameters has more noticeable downstream effects.

5.1 Chapters

At the chapter level, GraM-SUM outperforms the perplexity-based model for all chain lengths from 2 to 10. This difference is pronounced when the chain length is small; for $k \leq 5$, GraM-SUM scores at least one 1 ROUGE-1 point higher than the perplexity-based model. This suggests that GraM-SUM is better at identifying which paragraph-level summaries are most useful for constructing concise chapter-level summaries. For long chains, the difference between the models persists but is less obvious. This is expected since long chains will incorporate most, if not all, of the available paragraph-level summaries.

5.2 Books

Our results are similar at the book level. For short chains, GraM-SUM is superior to the perplexity model; at $k = 3$, GraM-SUMscores 5.90 ROUGE-1 points higher. This is in part due to a compounding effect that occurs during book-level summarization. Since the chapter-level summaries constructed by GraM-SUM generally score higher than those constructed by the perplexity model, the book-level summaries constructed from those

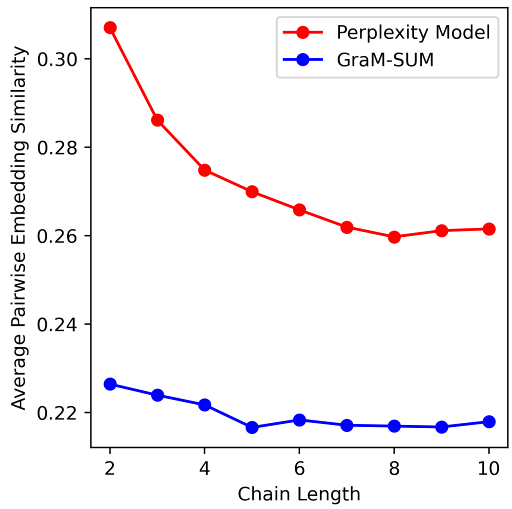

Figure 3: A comparison of the average pairwise embedding similarity of summaries produced by GraM-SUM and the perplexity model. For all chain lengths, GraM-SUM produces summaries with more diverse sentence-level embeddings.

chapter-level summaries will also tend to score higher.

For long chains, GraM-SUM is similar in performance (or even slightly worse) than the perplexity model. One explanation is that the diversity objective in GraM-SUM incentivizes chains with high recall. Initially, when k is small and there is little overlap between chain elements, the ROUGE recall and F_1 scores improve rapidly. As the chains get longer and more redundant, however, the recall stagnates; but because more text is added to each summary, the ROUGE precision and F_1 scores drop. Thus, compared to the perplexity model which does not explicitly optimize for recall, GraM-SUM reaches its optimal chain length more quickly. We do note that GraM-SUM's best ROUGE-1 score of 41.37 at $k = 7$ is higher than the perplexity model's best score of 40.82 at $k = 8$.

5.3 Embedding Diversity

In Figure 3, we demonstrate that the sentence-level embeddings for book summaries generated by GraM-SUM are more diverse than those of summaries produced by the perplexity model. By explicitly incentivizing diversity among chain elements, GraM-SUM is able to create richer high-level summaries.

5.4 Additional Observations

Despite its superior performance, GraM-SUM is significantly slower than the perplexity-based model. Whereas the run-time of the perplexity model on BookSum is on the order of minutes, GraM-SUM takes close to one hour to complete. In an offline setting, this difference can mostly be ignored, but it may be problematic in situations where low latency is preferred.

We also notice that GraM-SUM tends to favor longer low-level summaries when forming chains, while the perplexity model does the opposite. For $k = 10$, the average book-level summary produced by GraM-SUM is 1625 tokens long, compared to 1208 tokens for the perplexity model. It may be worthwhile to incorporate a length penalty into GraM-SUM's chain selection algorithm in order to discourage it from picking the longest available low-level summaries.

6 Conclusions

In this paper we introduced GraM-SUM, a graph-based model for summarizing book-length texts. The model operates at three levels of hierarchy: paragraphs, chapters, and books. At the paragraph level, GraM-SUM leverages a `t5-small` encoder-decoder transformer to generate new summaries. At the chapter and book levels, it uses an algorithm adapted from Gorinski and Lapata (2015) to select chains of low-level summaries that maximize progression, diversity, and importance.

We showed that for short chains, GraM-SUM outperforms the perplexity-based baseline on the BookSum dataset, and generally produces summaries with more diverse sentence embeddings. Overall, we believe GraM-SUM constitutes a step towards automatically producing short, coherent, and comprehensive summaries of long texts.

In future work, we should strive to incorporate larger encoder-decoder models to serve as the base GraM-SUM. As can be seen in the appendix, the paragraph-level summaries generated by `t5-small` are not always coherent when strung together. The model sometimes focuses on the wrong part of the text, which in turn causes the chapter-level and book-level summaries to seem disjointed. Using a more powerful transformer as the base for GraM-SUM may help it produce more eloquent summaries.

Other avenues for research include the addition of new graph-based features (e.g. social networks between entities), and the generation of more low-level summaries to produce a richer search space at the chapter and book levels. Finally, we should consider evaluating summaries on metrics other than ROUGE, in order to incentive summaries that are articulate as well as comprehensive.

References

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The long-document transformer.

Philip John Gorinski and Mirella Lapata. 2015. Movie script summarization as graph-based scene extraction. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1066–1076, Denver, Colorado. Association for Computational Linguistics.

Wojciech Kryściński, Nazneen Rajani, Divyansh Agarwal, Caiming Xiong, and Dragomir Radev. 2021. Booksum: A collection of datasets for long-form narrative summarization.

Rada Mihalcea. 2004. Graph-based ranking algorithms for sentence extraction, applied to text summarization. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 170–173, Barcelona, Spain. Association for Computational Linguistics.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.

Jeff Wu, Long Ouyang, Daniel M. Ziegler, Nisan Stiennon, Ryan Lowe, Jan Leike, and Paul F. Christiano. 2021. Recursively summarizing books with human feedback. *CoRR*, abs/2109.10862.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489, San Diego, California. Association for Computational Linguistics.

A Chapter-Level Examples

Below, we compare the chapter-level summaries produced by GraM-SUM and the perplexity model for the first chapter of *The Prince*.

A.1 GraM-SUM

Lodovico Moro is a son of Francesco Sforza, who married Beatrice d'Este, who ruled over Milan from 1494 to 1500, and died in 1510. Louis XII, King of France, born 1462, died 1515, and Charles VIII, king of France born 1470, died 1498. Louis was brought into Italy by the ambition of the Venetians, who wanted to obtain half the state of Lombardy by his intervention. Louis XII divorced his wife, Jeanne, daughter of Louis, and married Anne of Brittany, widow of Charles VIII, in order to retain the Duchy for the crown. He was Georges d'Amboise, created a cardinal by Alexander VI. Born 1460, died 1510. All states, all powers, that have held and hold rule over men have been and are either republics or principalities. The new are either entirely new, as was Milan to Francesco Sforza, or they are, as it were, members annexed to the hereditary state of the prince who has acquired them. Such dominions thus acquired are either accustomed to live under a prince, or to live in freedom; and are acquired either by the arms of the Prince himself or of others, or else by fortune or by ability. The Romans, in the countries that they annexed, observe closely these measures; they sent colonies and maintained friendly relations with the minor powers, without increasing their strength; they kept down the greater, and did not allow strong foreign powers to gain authority. They were kept friendly by them, the kingdom of Macedonia was humbled, Antiochus was driven out; yet the merits of the Achaeans and Aetolians never secured for them permission to increase their power.

A.2 Perplexity Model

Lodovico Moro is a son of Francesco Sforza, who married Beatrice d'Este, who ruled over Milan from 1494 to 1500, and died in 1510. Then Louis loses Lombardy by not having followed any of the conditions observed by those who have taken possession of countries and wish to retain them. Louis XII, King of France, born 1462, died 1515, and Charles VIII, king of France born 1470, died 1498. Louis was brought into Italy by the ambition of the Venetians, who wanted to obtain half the state of Lombardy by his intervention. The Romans, in the countries that they annexed, observe closely these measures; they sent colonies and maintained friendly relations with the minor powers, without increasing their strength; they kept down the greater, and did not allow strong foreign powers to gain authority. They were kept friendly by them, the kingdom of Macedonia was humbled, Antiochus was driven out; yet the merits of the Achaeans and Aetolians never secured for them permission to increase their power. He is forced to come into Italy, and he has to take an associate to the kingdom of Naples, where he is the prime arbiter in Italy.

B Book-Level Examples

Below, we compare the book-level summaries produced by GraM-SUM and the perplexity model for the full text of *The Prince*.

B.1 GraM-SUM

Lodovico Moro is a son of Francesco Sforza, who married Beatrice d'Este, who ruled over Milan from 1494 to 1500, and died in 1510. Louis XII, King of France, born 1462, died 1515, and Charles VIII, king of France born 1470, died 1498. Louis was brought into Italy by the ambition of the Venetians, who wanted to obtain half the state of Lombardy by his intervention. Louis XII divorced his wife, Jeanne, daughter of Louis, and married Anne of Brittany, widow of Charles VIII, in order to retain the Duchy for the crown. He was Georges d'Amboise, created a cardinal by Alexander VI. Born 1460, died 1510. All states, all powers, that have held and hold rule over men have been and are either republics or principalities. The new are either entirely new, as was Milan to Francesco Sforza, or they are, as it were, members annexed to the hereditary state of the prince who has acquired them. Such dominions thus acquired are either accustomed to live under a prince, or to live in freedom; and are acquired either by the arms of the Prince himself or of others, or else by fortune or by ability. The Romans, in the countries that they annexed, observe closely these measures; they sent colonies and maintained friendly relations with the minor powers, without increasing their strength; they kept down the greater, and did not allow strong foreign powers to gain authority. They were kept friendly by them, the kingdom of Macedonia was humbled, Antiochus was driven out; yet the merits of the Achaeans and Aetolians never secured for them permission to increase their power. Then, if you think about the nature of the government of Darius, you will find it similar to the kingdom of the Turk, and therefore it was only necessary for Alexander to overthrow him in the field, and then to take the country from him. The King of France is governed by one lord, the others are his servants, and dividing his kingdom into sanjaks, he sends there different administrators, and shifts and changes them as he chooses. He who considers both of these states will recognize great difficulties in seizing the state of the Turk, but, once it is conquered, great ease in holding it. Alexander the Sixth, in wishing to aggrandize the duke, his son, had many immediate and prospective difficulties. Firstly, he did not see his way to make him master of any state that was not a state of the Church; and if he was willing to rob the Church he knew that the Duke of Milan and the Venetians would not consent, because Faenza and Rimini were already under the protection of the Venices. The Duke of Milan, who died in 1466, is a natural daughter of Filippo Visconti. Machiavelli was the accredited agent of the Florentine Republic to Cesare Borgia (1478-1507) during the transactions which led up to the assassinations of the Orsini and Vitelli at Sinigalia, and along with his letters to his chiefs in Florence, he has left an account of the proceedings of the duke in his "Descritione del modo nello - a new prince. He weakened the Orsini and Colonnesi parties in Rome by gaining to himself all their adherents, making them his gentlemen, giving them good pay, and honouring them with office and command in such a way that in a few months all attachment to the factions was destroyed and turned entirely to the duke. Nabis, a tyrant of the Spartans, sustained the attack of all Greece and of a victorious Roman army, and against them he defended his country and his government, and for the overcoming of this peril it was only necessary for him to make himself secure against a few, but this would not have been sufficient had the people been hostile. The first method will be illustrated by two examples – one ancient, the other modern – and without entering further into the subject, he considers these two examples will suffice those who may be compelled to follow them. Agathocles, the Sicilian, becomes King of Syracuse from a private position, but he is a low and abject position. After the parricide, Oliverotto is strangled and strangled by Cesare Borgia, who took him with the Orsini and Vitelli at Sinigalia, which he had made his leader in valour and wickedness. He says that a prince should live amongst his people in such a way that no unexpected circumstances, whether of good or evil, shall make him change. The king of France has been able to drive him from Italy, and to ruin the Venetians, but he does not seem superfluous to recall it in some measure to memory. In Italy, the Pope, the Venetians, the Duke of Naples, and the Florentines have a lot of anxiety. Charles VIII invades Italy in 1494, and Alexander the Sixth arose afterwards, who of all the pontiffs that have ever been shown how a pope with both money and arms was able to prevail; and through the instrumentality of the Duke Valentino and by reason of the entry of the French, he brought about all those things which I have discussed above in the actions of the duke. The Pope finds the Church strong, possessing all the Romagna, the barons of Rome reduced to impotence, and, through the chastisements of Alexander, the factions wiped out; he also finds the way open to accumulate money in a manner such as had never been practised before Alexander's time. The Prince of Siena, Pandolfo Petrucci, ruled his state more by those who had been distrusted than by others. The prince always extracts more profit from them than from those who, serving him in too much security, may neglect his affairs. Antiochus goes to Greece, being sent for by the Aetolians to drive out the Romans. He sends envoys to the Achaeans, who were friends of the Roma, exhorting them to remain neutral, and on the other hand, the Roma urged them to take up arms. Maximilian I, born in 1459, died 1519, Emperor of the Holy Roman Empire. He married, first, Mary, daughter of Charles the Bold; after her death, Bianca Sforza; and thus became involved in Italian politics. Catherine Sforza, a daughter of Galeazzo and Lucrezia Landriani, born 1463, died 1509, and Nicolo Machiavelli was sent as envoy on 1499 to the countess of Forli. Then, he says, "The old Roman valour is not dead, Nor in th' Italians' brests extinguished."

B.2 Perplexity Model

The first case has been discussed, but we will speak of it again if it recurs. In the second case, one can say nothing except to encourage such princes to provision and fortify their towns, and not on any account to defend the country. They are free, they own but little country around them, and they yield obedience to the emperor when it suits them, because they are fortified in such a way that every one thinks the taking of them by assault would be tedious and difficult, seeing they have proper ditches and walls, they have sufficient artillery, and always keep in public depots enough for one year's eating, drinking, and firing. He says that a man who wants to be more highly placed is reputed liberal, another miserly, using a Tuscan term, and that all men are remarkable for some of those qualities which bring them either blame or praise. He says that if he maintains the name of liberal, he will be compelled to unduly weigh down his people, and tax them, and do everything he can to get money. It is wiser to have a reputation for meanness which brings reproach without hatred, rather than to be compelled through seeking an reputation for liberality to incur a name for rapacity which begets reproach with hatred. He says that the prince must be a violator of the property and women of his subjects, and that the majority of men live content, and he has only to contend with the ambition of a few, whom he can curb with ease in many ways. The prince is highly esteemed and is not easily conspired against, because he is defended by being well armed and having good allies. He says that whoever will consider it will acknowledge that either hatred or contempt has been fatal to the above-named emperors, and it will be recognized how it happened that, a number of them acting in one way and one in another, only one in each way came to a happy end and the rest to unhappy ones. He should choose the wise men in his state, and give them only the liberty of speaking the truth to him, and then only of those things of which he inquires, and of none others; but he ought to question them upon everything, and listen to their opinions, and afterwards form his own conclusions. He says that fortune is a woman, and if you wish to keep her under it is necessary to beat and ill-use her; and it is seen that she allows herself to be mastered by the adventurous rather than by those who go to work more coldly. This is a great opportunity to let Italy see her liberator appear, but it is not possible to express the love with which he would be received in all those provinces which have suffered so much from these foreign scourings. Catherine Sforza, a daughter of Galeazzo and Lucrezia Landriani, born 1463, died 1509, and Nicolo Machiavelli was sent as envoy on 1499 to the countess of Forli. Then, he says, "The old Roman valour is not dead, Nor in th' Italians' brests extinguished."